

Sapir Academic College

From the Selected Works of Prof. Boaz Sangero

Fall September 5, 2018

Safety from Flawed Forensic Sciences Evidence

Boaz Sangero



Available at: https://works.bepress.com/dr_boaz_sangero/14/

August 2018

Safety from Flawed Forensic Sciences Evidence

Boaz Sangero

Academic Center of Law & Business, Israel, sangero@clb.ac.il

Follow this and additional works at: <https://readingroom.law.gsu.edu/gsulr>

 Part of the [Administrative Law Commons](#), [Comparative and Foreign Law Commons](#), [Constitutional Law Commons](#), [Courts Commons](#), [Criminal Law Commons](#), [Criminal Procedure Commons](#), [Disaster Law Commons](#), [Evidence Commons](#), [International Law Commons](#), [Law and Society Commons](#), [Law Enforcement and Corrections Commons](#), [Legal Ethics and Professional Responsibility Commons](#), [Public Law and Legal Theory Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

Boaz Sangero, *Safety from Flawed Forensic Sciences Evidence*, 34 GA. ST. U. L. REV. 1129 (2018).

Available at: <https://readingroom.law.gsu.edu/gsulr/vol34/iss4/8>

This Article is brought to you for free and open access by the Publications at Reading Room. It has been accepted for inclusion in Georgia State University Law Review by an authorized editor of Reading Room. For more information, please contact mbutler@gsu.edu.

SAFETY FROM FLAWED FORENSIC SCIENCES EVIDENCE

Boaz Sangero*

I. Introduction—Modern Safety

This article addresses the way to safety in the context of forensic sciences evidence. After presenting the current lack of safety, which I term “unsafety,” I raise some possible safety measures to contend with this. My suggestions are grounded on two bases: first, the specific analysis of each type of evidence in line with the most recent research on the subject; and second, modern safety theory and its application to the criminal justice system. It is important to stress that my proposals represent only some of the conceivable safety measures. Developing a comprehensive safety theory for the criminal justice system will require considerable additional cross-disciplinary research work, which I recommend be undertaken within the framework of a Safety in the Criminal Justice System Institute (SCJSI).¹

I have chosen, for discussion purposes, to analyze the two central types of forensic sciences evidence currently predominating criminal law: DNA testing and fingerprint comparisons. For each of these, I will review the most up-to-date research on the topic. I will demonstrate why present use of these types of evidence is not error-free and fails to ensure safety from false convictions and then offer

* Professor of Law, Founder of the Criminal Law & Criminology Department, College of Law & Business, Ramat-Gan, Israel; and School of Law, Sapir College, Israel. I thank Prof. Rinat Kitai-Sangero for her tremendous help in all the stages of the writing.

1. Introducing modern safety into systems lacking a culture of safety requires the establishment of a special institute to carry out this function, and the securing of resources necessary for the new institute to operate in a meaningful way. Thus, for example, in the field of aviation, the Federal Aviation Administration (FAA) was established; in the field of transportation, the National Transportation Board (NTSB) was founded; in the area of food and drugs, there is the Food and Drug Administration (FDA); the Occupational Safety and Health Administration (OSHA) serves the occupational field; and various such bodies were established in the medical field, such as the National Center for Patient Safety (NCPS) and the Center for Patient Safety Research and Practice (CPSRP). In all of these fields, the recognition of safety issues and the need to improve performance led to national focus on safety leadership, the development of a knowledge base, and the distribution of information—an agenda to which substantial resources were devoted.

different ways of improving safety in these contexts. I will propose a general solution regarding all types of forensic evidence, based on an earlier proposition I developed with Dr. Mordechai Halpert, namely, that the legislature must enact a rule that precludes the admissibility of forensic evidence in court unless it has been developed as a “safety-critical system.”² The knowledge and solutions for developing safety-critical devices already exists in other engineering fields, such as medical devices and aviation devices. Thus, all that is needed is the willingness and reasonable resources to carry this out. Later in this article, I will offer some additional general solutions.

This article also addresses what is known as “junk science,” which refers to evidence that is presented, inaccurately and misleadingly, as scientific evidence when it has, at best, a flimsy connection to science. Despite studies clearly pointing out this lack of scientific grounding, including the 2009 National Academy of Sciences (NAS) Report,³ some courts still admit such evidence as scientific evidence. This is proof in itself of just how far the criminal justice system is from being a safe system.

In certain fields, the meaning of a “safety-critical system” is well understood, and resources are, therefore, invested in modern safety methods that significantly reduce the rate of accidents. This is the case, for example, in the field of pharmaceuticals and drugs, where in the first half of the twentieth century the need for safety was already acknowledged and internalized and the necessary powers and authorities were granted to the FDA to ensure this. This was also the case in the aviation field, which abandoned the obsolete “Fly-Fix-Fly” approach in the mid-twentieth century and developed more advanced safety methods that generally follow an “Identify-Analyze-Control” model and are aimed at “First-Time-Safe.” The latter

2. Mordechai Halpert & Boaz Sangero, *From a Plane Crash to the Conviction of an Innocent Person: Why Forensic Science Evidence Should Be Inadmissible Unless It Has Been Developed as a Safety-Critical System*, 32 *HAMLIN L. REV.* 65, 93–94 (2009).

3. NAT’L RESEARCH COUNCIL, COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIS. CMTY. ET AL., *STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD* (Feb. 18, 2009) [hereinafter *NAS-2009 Report*].

approach involves systematic identification of future hazards, analysis of the probability of their occurrence, and complete neutralization of the risk, or at least its reduction, to an acceptable level. Modern safety approaches such as these were implemented in other fields as well, such as transportation and engineering, and later on, in labor and medicine. These safety systems are constructed on, among other things, safety education and training; a culture of safety; a duty to report not only accidents but also incidents (near-accidents); professional risk assessment; a process of perpetual improvement; and the understanding that safety in each component of a system alone in detachment from the entire system is not sufficient for achieving system safety.

Accidents also happen in the criminal justice system, of course, in the form of false convictions. For this reason, this system must also be classified as a safety-critical system. As systems of this type entail matters of life and death, any system error is likely to cause severe harm to both individuals and society at large. A false conviction is a system error and accident just like a plane crash, not only from a metaphorical perspective but also in the very realistic terms of economic cost.⁴ Yet, in criminal law, a “Hidden Accidents Principle” governs.⁵ Thus, the overwhelming majority of false convictions are never detected, which leads to the erroneous traditional and conservative assumption that they occur at an almost negligible rate and that the criminal justice system is “almost” perfect. Consequently, little thought has ever been given to safety in the system, and therefore the criminal justice system, from a safety perspective, lags far behind other areas of life.

4. Boaz Sangero & Mordechai Halpert, *A Safety Doctrine for the Criminal Justice System*, MICH. ST. L. REV. 1293, 1304–05 (2011). The idea of incorporating into the criminal justice system a modern safety theory that is commonly accepted in other areas—such as space, aviation, engineering, and transportation—was developed jointly by myself and Dr. Mordechai Halpert and presented in a number of coauthored articles, particularly *A Safety Doctrine for the Criminal Justice System*, *id.* My current article is intended to expand on the preliminary proposition and engage in the application of the modern safety theory in the criminal justice system, specifically regarding forensic sciences evidence. Halpert & Sangero, *supra* note 2, at 93–94.

5. Sangero & Halpert, *supra* note 4, at 1314–16.

The patently flawed assumption of a low false-conviction rate has been challenged in recent decades, primarily because of the work of the Innocence Project. The Project exposes hundreds of cases of false convictions through genetic testing and empiric studies based on the Project's findings, which point to a very high false-conviction rate: at least 5% for the most serious crimes and apparently an even higher rate for less serious crimes.⁶ About half of those false convictions were based on (presumably false) forensic sciences evidence.⁷

This Article proceeds as follows. Part II demonstrates the current state of unsafety in forensic sciences evidence. It starts from a discussion of the admissibility of forensic sciences evidence, from *Frye* to *Daubert*. It then moves to a discussion of the weight of forensic sciences evidence, from “uniqueness,” “individualization,” and “perfection” to empirical and probabilistic foundation. It ends with showing why a single piece of evidence should not be sufficient for a conviction, using Bayes' theorem and medical diagnostics. Then the two central types of forensic sciences evidence currently predominating criminal law—DNA testing and fingerprint comparisons—are discussed at length. Part II ends with a critique of “junk science” as evidence. Part III suggests safety measures.

II. *Unsafety in Forensic Sciences Evidence*

A. *The Admissibility of Forensic Sciences Evidence: From Frye to Daubert*

How can judges distinguish between “true” scientific evidence and inadmissible “junk” scientific evidence? In the past, the *Frye* standard, set in 1923, was the prevailing rule in American law.⁸ Under this rule, “general acceptance” of the method on which the

6. D. Michael Risinger, *Innocents Convicted: An Empirically Justified Factual Wrongful Conviction Rate*, 97 J. CRIM. L. & CRIMINOLOGY 761, 786–88 (2007).

7. BRANDON L. GARRETT, CONVICTING THE INNOCENT—WHERE CRIMINAL PROSECUTIONS GO WRONG 279 fig. A.5 (2011).

8. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923); BOAZ SANGERO, CONVICTING THE INNOCENT IN ISRAEL AND IN OTHER COUNTRIES—CAUSES AND SOLUTIONS 129 (2014) (Isr.).

evidence rests in the relevant scientific field is sufficient for it to be admissible in court.⁹ This rule predominated for seventy years until it was supplanted by a more sophisticated rule, set in 1993 by *Daubert*.¹⁰ Whereas the *Frye* rule enabled judges to refrain from a deep examination of the relevant scientific field and suffice with the fact that the method in question is generally accepted by the scientific community to which it belongs, the *Daubert* rule is designed to ensure greater caution: it made judges “gatekeepers,” tasked with blocking the entry of nonscientific evidence fraudulently presented as scientific into the courtroom.

Daubert set four admissibility criteria for scientific evidence.¹¹ First, the scientific method on which the evidence is allegedly based must adhere to Karl Popper’s principle of falsifiability. That is, the method must be empirically testable for falsifiability or refutability, and must have been successfully tested in order to be considered scientific methodology.¹² Second, the error rate of the evidence must be known.¹³ Third, the scientific method must have been subject to peer review and published.¹⁴ And finally, similar to the requirement under the *Frye* rule, the method must be accepted within the relevant scientific community.¹⁵ As Professor Alex Stein compellingly explained, the fourth requirement is particularly crucial: admitting scientific evidence that does not meet the *Frye* standard would undermine the allocation of the risk of error, which cannot be contingent on a judicial prediction regarding the odds of the success of a particular scientific progress.¹⁶

9. *Frye*, 293 F. at 1014.

10. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 592–94 (1993).

11. *Id.*; Boaz Sangero & Mordechai Halpert, *Scientific Evidence v. “Junk Science,”* 11 C.L.B. L. STUD. 425, 430 (2014) (Isr.); Kristen Bolden, *DNA Fabrication, a Wake Up Call: The Need to Reevaluate the Admissibility and Reliability of DNA Evidence*, 27 GA. ST. U. L. REV. 409, 419–24 (2011).

12. *Daubert*, 509 U.S. at 593.

13. *Id.* at 594.

14. *Id.* at 593.

15. *Id.* at 594.

16. Alex Stein, *Against Free Proof*, 31 ISR. L. REV. 573, 587 (1997).

The *Daubert* standard, along with certain refinements of the rule, thus requires judges to make an effort to delve into the relevant scientific field and to examine the reliability and validity of the scientific method (which is, in effect, a classification system) on which the evidence is grounded.¹⁷ “Reliability” in this context refers to the consistency of the classification; for example, a test is reliable if the same result will be arrived at if performed by different experts at different points in time. “Valid” means that the method (the classification system) is suited to the purposes for which it is used. In the legal context, a test is valid if it suits the aim of distinguishing between who is guilty of committing the crime in question and who is innocent. A test can be reliable but not valid. For example, a test that uses an arbitrary rule to distinguish between guilt and innocence according to skin color will be a reliable test, because different people will concur regarding the color of a defendant’s skin and color will not alter over time. This method will not be valid, though, because the color of a person’s skin is immaterial to the question of his guilt or innocence. In contrast, reliability is a necessary condition for validity: if different experts can arrive at different outcomes, or if outcomes vary over time, then the test or method cannot be valid.

In effect, then, despite the fact that the *Daubert* rule takes precedence today, the two rules—*Daubert* and *Frye*—exist side-by-side in American case law.¹⁸ Generally, emphasis on the criterion of acceptance by the relevant scientific community (the *Frye* test) will lead the court to relax the requirements for recognizing the scientific reliability of given evidence, whereas a deeper examination of the scientific grounding and methodology of the evidence—in the spirit

17. Sangero & Halpert, *supra* note 11, at 431. On validity and reliability in the legal context, see Sandy A. Zabell, *Fingerprint Evidence*, 13 J.L. & POL’Y 143, 154 (2005).

18. Bolden, *supra* note 11, at 419–24; Sangero & Halpert, *supra* note 11, at 431. The *Daubert* standard has been adopted in federal courts, where the federal rules of evidence also apply and set a similar arrangement. Fed. R. Evid. 702–03. For a review of the various states that have adopted the *Daubert* rule as opposed to those that have continued to follow the *Frye* rule, see Edward K. Cheng & Albert H. Yoon, *Does Frye or Daubert Matter? A Study of Scientific Admissibility Standards*, 91 VA. L. REV. 471, 472–73 (2005).

of the *Daubert* rule—will lead to a more rigorous approach.¹⁹ Oftentimes, as judges struggle with the investigation of the scientific aspects with which the *Daubert* rule is concerned, they incorrectly classify methodological errors made by experts as relevant to the weight—and not to the actual validity—of the evidence, admitting evidence as scientific when it is not deserving of the title, at times through resort to the *Frye* rule.²⁰

One might wonder why scientific evidence in particular should be required to meet the criteria of reliability and validity when other types of evidence, such as witness testimony, are not. There are several possible interrelated, cumulative answers to this.²¹ First, scientific evidence is purported to be science (expert testimony is an exception to the hearsay rule). Second, the “aura” of science can be expected to blind judges and jurors and lead them to overestimate the real probative strength of scientific evidence. Third, in scientific evidence, it is more practical to require and examine reliability and validity. And finally, fourth, the path toward implementing appropriately rigorous standards for types of evidence that can determine defendants’ fate must begin somewhere. The next step is to address the remaining types of evidence and apply similar rules and standards to them.

B. The Weight of Forensic Sciences Evidence: From “Uniqueness,” “Individualization,” and “Perfection” to Empirical and Probabilistic Foundation

Presumably, the different types of scientific evidence could have been expected to greatly advance and refine criminal law and make it more precise. There is no doubt that DNA testing and fingerprint comparisons—despite being termed “circumstantial” evidence—are

19. David L. Faigman, *Admissibility Regimes: The “Opinion Rule” and Other Oddities and Exceptions to Scientific Evidence, the Scientific Revolution, and Common Sense*, 36 SW. U. L. REV. 699, 701–703 (2008).

20. Jay P. Kesan, *An Autopsy of Scientific Evidence in a Post-Daubert World*, 84 GEO. L.J. 1985, 2040 (1996).

21. Sangero & Halpert, *supra* note 11, at 432.

far more accurate than the traditional types of evidence, referred to as “direct” evidence: eyewitness testimony and confessions.²² Yet in the framework of the first Innocence Project, misapplication of forensic science is the second most common contributing factor to wrongful convictions, found in nearly half (45%) of DNA exoneration cases.²³ Moreover, as of March 2018, in the National Registry of Exonerations, 517 of the 2,152 registered exonerations (24%) involved false or misleading forensic evidence.²⁴

Why is it that accurate forms of evidence can be more misleading than other types of evidence? One possible explanation is that they blind judges and jurors. As one defense attorney remarked, “If you put God on the witness stand . . . and God’s testimony conflicted with the DNA evidence, everyone would automatically say, ‘Why is God lying like this?’”²⁵ We all, justifiably, hold science in the highest esteem. However, judges and jurors are not scientists. Judges are trained jurists who are required to reach determinations on a wide variety of issues, from all spheres of life, some of which are from the scientific field. In present times, human knowledge is so vast that no one has the capacity to be an expert in a number of fields, to be a “Renaissance person.” Accordingly, judges must rely on experts who present them with the results of scientific tests they performed. As it is the judges and jurors—and not the experts—who are the triers of fact, judges can and should demand experts not simply testify about the results of the tests they performed, but describe in detail their methodology for judges and jurors to examine. This, of course, is

22. Jennifer L. Mnookin, *Fingerprints: Not a Gold Standard*, 20 ISSUES IN SCI. & TECH. 47, 53 (2003).

23. *Misapplication of Forensic Science*, INNOCENCE PROJECT, <https://www.innocenceproject.org/causes/misapplication-forensic-science/> [<https://perma.cc/4JNC-PAVN>] (last visited July 7, 2018).

24. NAT’L REGISTRY OF EXONERATIONS, EXONERATIONS BY CONTRIBUTING FACTOR, <https://www.law.umich.edu/special/exoneration/Pages/ExonerationsContribFactorsByCrime.aspx> [<https://perma.cc/9HXC-L4LT>] (last visited July 7, 2018).

25. J. Herbie DiFonzo, *The Crimes of Crime Labs*, 34 HOFSTRA L. REV. 1, 1 (2005) (citing Laura Lafay, *Reasonable Doubt*, STYLE WKLY. (July 6, 2005) <https://www.styleweekly.com/richmond/reasonable-doubt/Content?oid=1380129> [<https://perma.cc/2YU8-6YT2>] (quoting defense attorney David Baugh)).

contingent on proof of the validity and reliability of the scientific method used by the expert. Judges should require experts produce scientific written support of their approaches, but not all judges do so. Many rely on the experts almost blindly, to the point where, in practice, they serve as no more than a rubber stamp.²⁶

In addition, as the law is based on legal precedents, if a court—especially a higher appellate court—makes the mistake of incorrectly admitting a specific type of evidence as “scientific,” it likely sets a precedent for other courts, which will follow the precedent without re-examining the scientific reliability of the method. In this way, junk science, too, could be found admissible by one court, and other courts would follow suit without engaging in the necessary scrutiny; this is precisely what happened over the course of many years with microscopical hair comparisons, footprint comparisons, and voice comparisons. Furthermore, even when (genuine) scientists find in their research that certain allegedly “scientific” types of evidence are not grounded in science and are unreliable and invalid, many judges, who are used to basing convictions on such evidence, have difficulty accepting this as it would mean conceding their own past mistakes.²⁷

In practice, a random match is possible with every type of test (though with DNA testing, the probability of such a match is extremely slim). With all tests (including DNA testing), moreover, there is a real possibility of a lab error or an error in the expert’s interpretation of the results. In contrast to the prevailing perception of the objectivity of all determinations in expert testimony on scientific evidence, certain issues are in fact contingent on the expert’s subjective interpretation of the findings.²⁸ And indeed, the chance of error, which always exists, combined with the fallacy of the transposed conditional has been the underlying cause of many false convictions.²⁹

26. SANGERO, *supra* note 8, at 105–06.

27. D. Michael Risinger, *Whose Fault?—Daubert, the NAS Report, and the Notion of Error in Forensic Science*, 38 *FORDHAM URB. L.J.* 519, 539 (2010); Sangero & Halpert, *supra* note 11, at 443.

28. Sangero & Halpert, *supra* note 11, at 443.

29. *See infra* Section II.C.

In addition, forensic lab staff work very closely with the police and prosecution and are often even directly subordinate to them. Police investigators supply lab workers with investigation details that are completely irrelevant to the required lab testing, but which strengthen the conception of the suspect's guilt, such as the fact that a suspect confessed or was identified by a witness. Information of this sort is likely to bias the test results.³⁰

Jennifer Mnookin described the current situation well:

Forensic scientists have regularly testified in courts to matters that are, quite honestly, both less proven and less certain than they are claimed to be. They have overstated their degree of knowledge, underreported the chances of error, and suggested greater certainty than is warranted. More generally, many kinds of forensic science are not entirely based on methods and approaches that we usually associate with validated research science. Their claims and the limits to their claims are not closely based on or constrained by the formal collection of data. Their empirical assertions are not grounded in careful research that has been subject to peer review and publication. There has been remarkably little formal validation of their methods. And there has been far too little study of how often forensic scientists might make mistakes, and when or why these possible errors are more likely to occur.³¹

From the perspective of the historical development of evidence law, every generation has realized that the weight once accorded to certain types of evidence was excessive.³² Over the years, evidence that was formerly held to be very strong emerges, in both the research and in

30. See Jennifer L. Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 BROOK. L. REV. 1209, 1215 (2010); DiFonzo, *supra* note 25, at 4; William C. Thompson, *Beyond Bad Apples: Analyzing the Role of Forensic Science in Wrongful Convictions*, 37 SW. L. REV. 1027, 1034 (2008).

31. Mnookin, *supra* note 30, at 1210.

32. SANGERO, *supra* note 8, at 47.

practice, as not as accurate as thought. Thus, for example, we know today that certain types of direct evidence once considered “classic” evidence—eyewitness testimony and confessions to the police³³—are problematic and have led to many false convictions.³⁴ During the past century, scientific evidence has become the evidentiary “false messiah.” To generalize, forensic scientists present scientific evidence in court as unequivocal evidence that reflects the uniqueness of every individual from all other human beings and produces a perfect identification of every person. Yet this premise of uniqueness is nothing more than pure conjecture, with no data to support it; it is based on the assumption that nature (or God) never repeats itself. On the one hand, it exempts forensic scientists from the rigors of methodological research, data collection, and incidence calculation, and on the other hand, it allows them to assert purportedly certain identification in court. As Michael J. Saks and Jonathan L. Koehler clarified in an article published in the leading journal *Science*, “[T]he time is ripe for the traditional forensic sciences to replace antiquated assumptions of uniqueness and perfection with a more defensible empirical and probabilistic foundation.”³⁵ Moreover, they further observed elsewhere, “The concept of individualization, which lies at the core of numerous forensic science subfields, exists only in a metaphysical or rhetorical sense. There is no scientific basis for the individualization claims in forensic sciences.”³⁶

33. The confession has even been deemed “the queen of evidence,” rather than the “empress of false convictions” as it should be. See Boaz Sangero, *Miranda Is Not Enough: A New Justification for Demanding “Strong Corroboration” to a Confession*, 28 CARDOZO L. REV. 2794, 2894 n.11 (2007).

34. Boaz Sangero & Mordechai Halpert, *Proposal to Reverse the View of a Confession: From Key Evidence Requiring Corroboration to Corroboration for Key Evidence*, 44 U. MICH. J. L. REFORM 511, 516 (2011).

35. Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 895 (2005) [hereinafter Saks & Koehler, *Forensic Identification Science*]. For a more detailed discussion, see Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199 (2008) [hereinafter Saks & Koehler, *The Individualization Fallacy*].

36. Saks & Koehler, *The Individualization Fallacy*, *supra* note 35, at 202.

In a later article, Koehler and Saks summarized the issue as follows:

- (1) the data necessary to achieve individualization have never been collected for any of the forensic science fields which aspire to individualize the source of crime scene evidence to its sole possible contributor;
- (2) the best available—and perhaps the only scientifically defensible—approach to forensic identification is the use of random match probability estimates (which are not yet employed by any of the traditional forensic identification sciences);
- (3) the argument that all objects are discernibly unique stands on little more than an oft-repeated maxim of forensic science legend and the illusory intuition that small frequencies imply uniqueness;
- (4) probability estimates (by definition) cannot lead to uniqueness or individualization;
- (5) assertions of individualization generally exaggerate what is known or can be accomplished by forensic examiners.³⁷

Thus, it emerges that with the exception of the context of DNA comparisons, forensic experts tend to systematically violate the most basic scientific principle that probabilistic estimates (as well as verbal quantitative estimates) must be grounded in data.³⁸ As Koehler noted, “the specific language used in court by experts can be the difference between testimony that is truly helpful and testimony that

37. See Jonathan J. Koehler & Michael J. Saks, *Individualization Claims in Forensic Science: Still Unwarranted*, 75 BROOK. L. REV. 1187, 1187–88 (2010); but see David H. Kaye, *Probability, Individualization and Uniqueness in Forensic Science Evidence: Listening to the Academies*, 75 BROOK. L. REV. 1163, 1166 (2010).

38. Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1, 19 (2009); Sangero & Halpert, *supra* note 11, at 445; Mnookin, *supra* note 30, at 1210.

is confusing or unhelpful.”³⁹ Accordingly, he proposed that the forensic linguistic community should identify clear and consistent standards for reporting and testifying about test results, and that its members be trained in elementary statistics and probability.⁴⁰ In sum, Koehler and Saks pointedly asserted that forensic scientists should not be allowed “to say, in effect, ‘trust me: that’s the source.’ Real scientists don’t say ‘trust me.’ They provide data.”⁴¹

A scientific approach, however, should be taken not only with regard to the random match probability, but also to the possibility of errors in lab testing. Where the possibility of a random match is very rare (in genetic comparisons, for example, there is often a one-in-millions or even billions chance) and where the possibility of an error in the lab testing is far more common (at least one percentage rate in all likelihood), the courts and legislature must require that the prosecution provide also, and even principally, the latter figure as a condition for the admissibility of the evidence in court. This information is very important not only with respect to admissibility but also in terms of the correct weight to be attributed to the evidence. Currently, there is a tendency to ignore the possibility of error in lab testing, and consequently, scientific evidence is accorded far greater weight than what is appropriate.⁴² The possibility of a lab testing error is an inherent aspect of every scientific test. It appears that the law is perhaps the only sphere of life in which the tendency is to ignore this possibility. In the medical diagnostics field, for example, the incidence of the disease in the risk group to which the testee belongs is vital to correctly calculate the positive predictive value of the test and not fall victim to the “base rate fallacy.”⁴³ The

39. Jonathan J. Koehler, *Linguistic Confusion in Court: Evidence from the Forensic Sciences*, 21 J. L. & POL’Y 515, 515 (2013) (emphasis omitted).

40. *Id.* at 537–38.

41. Koehler & Saks, *supra* note 37, at 1208.

42. See, e.g., Simon A. Cole, *More than Zero: Accounting for Error in Latent Print Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985, 1034 (2005).

43. See, e.g., Stuart Spitalnic, *Test Properties 2: Likelihood Ratios, Bayes’ Formula, and Receiver Operating Characteristic Curves*, HOSP. PHYSICIAN, Oct. 2004, at 53, 53–54; see also *infra* Section II. C.

FDA even requires manufacturers of medical diagnostic devices to provide the positive predictive value of the test they manufacture for different levels of the incidence of a disease among different risk groups.⁴⁴

In the past, a common erroneous claim (referred to in the literature as “case specific” or “false positive fallacy”)⁴⁵ was that a court presumably has no need for statistics on laboratory error rates because it allegedly has the ability to check that all necessary procedures were followed in the specific case before it and, thereby, ensure that no error could have occurred.⁴⁶ Today, it is generally understood that courts are incapable of making sure that no error occurred, and that even a negligible probability of lab error should, in certain circumstances, result in a significant diminishment of the weight of the relevant evidence.⁴⁷ Errors occur even in the best of laboratories and even when the experts testify that all test protocols were followed.⁴⁸

The foremost authority in scientific evidence is the 2009 NAS Report, *Strengthening Forensic Science in the United States: A Path Forward*,⁴⁹ which was written by an interdisciplinary panel of distinguished scholars and practitioners. The report determined that the forensic science system does not function properly and that a significant improvement is required.⁵⁰ Despite the fact that forensic

44. Steven Gutman, *Food and Drug Administration (FDA)'s Impact on Laboratory Performance: FDA's Perspective*, 42 CLINICAL CHEMISTRY 786, 787 (1996); see also FOOD & DRUG ADMIN., STATISTICAL GUIDANCE ON REPORTING RESULTS FROM STUDIES EVALUATING DIAGNOSTIC TESTS (Mar. 13, 2007); Sangero & Halpert, *supra* note 11, at 445–46.

45. William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 J. FORENSIC SCI. 47, 51–52 (2003); see also *infra* Section II.C.

46. NAT'L RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 85 (1996) [hereinafter NRC-II Report].

47. NAS-2009 Report, *supra* note 3, at 135. “However, even a very small (but nonzero) probability of false positive can affect the odds that a suspect is the source of a sample with a matching DNA profile.” *Id.* at 130; see also *id.* at 121 (discussing the fact that no distinction should be made between a lab error and a random match probability error).

48. William C. Thompson, *Tarnish on the “Gold Standard”: Understanding Recent Problems in Forensic DNA Testing*, 30 CHAMPION 10, 11–12, 13–14 (2006); Sangero & Halpert, *supra* note 11, at 446.

49. NAS-2009 Report, *supra* note 3, at i; Sangero & Halpert, *supra* note 11, at 446.

50. NAS-2009 Report, *supra* note 3, at 14–19.

lab workers purport to engage in science, many lack the required training to do so, and there is an inadequate connection between their work and academic knowledge.⁵¹ This has led to significant knowledge gaps between scientists and forensic “scientists.”⁵² For the forensic science field to properly service the law, the organizational systems and structures must be significantly upgraded, training and qualifications improved, best practices adopted, and accreditation programs for laboratories made mandatory. The report emphasized that the most fundamental shortcoming emerged in the scientific knowledge of those who work in forensic science, and it called for genuine academic research in the area of forensic science to be conducted at universities.⁵³ Accordingly, the report recommended the establishment of an independent federal agency equipped with the necessary powers, authority, and resources to implement fundamental changes in the area of scientific evidence.⁵⁴ Obviously, such an agency would have very important ramifications for safety.

The report also recommended removing public forensic science laboratories from the administrative control of law enforcement agencies and prosecutors’ offices, standardizing laboratory reports, and instituting mandatory accreditation for all forensic laboratories, mandatory certification for all forensic workers, and a mandatory code of ethics.⁵⁵ Another very important recommendation was to conduct foundational research to assess the validity and reliability of methods used in the analysis of evidence, especially pattern evidence.⁵⁶

Of particular importance to our purposes are the report chapters that deal individually with each of the specific types of scientific evidence. Even putting aside any possible errors on the part of

51. *Id.* at 15.

52. *Id.*; see also Saks & Koehler, *The Individualization Fallacy*, *supra* note 35, at 215.

53. NAS-2009 Report, *supra* note 3, at 27–28.

54. *Id.* at 19; Sangero & Halpert, *supra* note 11, at 446–47.

55. NAS-2009 Report, *supra* note 3, at 24–26.

56. *Id.* at 22–23.

experts, it emerges that a considerable amount of “scientific evidence” is not grounded in science at all. The report estimated the scientific basis for each of the types of scientific evidence and types of pseudoscientific evidence, beginning with genetic comparisons—the most scientific type of evidence—and ending with junk science, such as the microscopical hair comparisons.⁵⁷ In the following individual discussions of some of these types of evidence, I will discard the “uniqueness,” “individualization,” and “perfection” myths and instead present their realistic weights in light of the NAS Report’s estimates.⁵⁸

C. Why a Single Piece of (Scientific) Evidence Should Not Be Sufficient for a Conviction: Bayes’ Theorem and Medical Diagnostics

1. General

Under the Hidden Accidents Principle in criminal law, most false convictions go undetected. This means that the way to safety in the criminal justice system must be found by drawing on the insights and experience of other fields, where accidents are observable. The medical diagnostics field can be particularly enlightening as to how to contend with the challenge of hidden accidents and flaws in the system,⁵⁹ as it is prominently characterized by its efforts to find ways to diagnose rare diseases in low-risk populations, which are hard to detect.⁶⁰

Indeed, using a single piece of evidence to determine the identity of the perpetrator of a crime can be analogized to using a single test to diagnose a rare disease. Just as a medical doctor should not base her diagnosis on a lone test without considering the statistical

57. *Id.* at 161.

58. Sangero & Halpert, *supra* note 11, at 447.

59. Boaz Sangero & Mordechai Halpert, *Why a Conviction Should Not Be Based on a Single Piece of Evidence: A Proposal for Reform*, 48 JURIMETRICS J. 43, 51–52, 90–94 (2007).

60. Klemens B. Meyer & Stephen G. Pauker, *Screening for HIV: Can We Afford the False Positive Rate?*, 317 NEW ENG. J. MED. 238, 239 (1987); Sangero & Halpert, *supra* note 34, at 529.

implications, law enforcement agents must be aware of the limitations of a single piece of evidence as an indicator of guilt.

2. *Bayes' Theorem As Odds*

Before I describe the Bayes' Theorem and the medical diagnostic model, the following example can illustrate why they are so critical.⁶¹ Assume that the manufacturer of a home HIV testing kit reports an average 0.1% false positive rate. Thus, if 10,000 non-carriers test themselves with this kit, 10 false positive HIV results would be obtained. Now, let us assume that John uses the kit to test himself and gets the positive result that he is an HIV carrier. What is the probability that he is truly a carrier? The obvious answer seems to be 99.9%, with only a 0.1% likelihood of a false positive. However, a crucial distinction should be made between conditional probability and inverse conditional probability: although the probability of a positive test result for a healthy person is indeed 0.1%, of relevance to us is the probability, given a positive test result, of the person tested actually being a carrier.

To illustrate, further assume that John is in a low-risk group for HIV: he practices safe sex; he does not use intravenous needles; he has never been given a blood transfusion. Say that the HIV incidence rate for this group is 1-in-10,000, which, in medical statistical terms, is the base rate (incidence of the disease) for the group.⁶² Thus, if 10,000 people in John's low-risk group were to test themselves using the home HIV testing kit, eleven would get a positive result: ten cases of error (false positives) for non-carriers ($9,999 \times 0.1\% = \sim 10$), and one case of an actual HIV carrier (because 1-in-10,000 members of this group is a carrier).⁶³ Consequently, contrary to what most intuitively presume, there is only a 1-in-11 (approximately 9%) likelihood that John is an HIV-carrier if he gets a positive test result,

61. This is a modified version of an example developed in Sangero & Halpert, *supra* note 34, at 529–532.

62. *Id.* at 529.

63. Assume that there is no possibility of a false negative—that is, that there will never be a negative test result for a carrier.

or, in other words, there is a probability of 10-in-11—about 91%—that this is a false positive.⁶⁴

This demonstrates how analyzing test results is not an intuitive matter. A probabilistic analysis that considers the incidence of the disease within the tested population will tend to show that a test that, at first glance, appears to have precise results is in fact completely inconclusive on its own. This failure to factor in the incidence of a disease is a cognitive failure known in the psychological literature as the “base-rate fallacy”⁶⁵ or “base-rate neglect”⁶⁶ and can be overcome through probabilistic analysis.⁶⁷

Bayes’ Theorem, which originated in the eighteenth century,⁶⁸ is very significant in applied probability theory, and can be expressed in the form of odds:

$$\text{Posterior Odds} = \text{Likelihood Ratio} \times \text{Prior Odds}^{69}$$

This simple theorem holds that by updating our initial belief about something with objective new information, we arrive at a new and improved belief:⁷⁰

A simple statement of Bayes’ Theorem uses three terms. One is the *prior odds* of a proposition—that is, the odds as assessed before receipt of the new evidence. The second is the *posterior odds* of the proposition—that is, the odds that the proposition is true as assessed after receipt of the new

64. Sangero & Halpert, *supra* note 34, at 530.

65. Amos Tversky & Daniel Kahneman, *Evidential Impact of Base Rates*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 153, 154 (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982).

66. Maya Bar-Hillel, *The Base-Rate Fallacy in Probability Judgments*, 44 ACTA PSYCHOLOGICA 211, 211 (1980); Tversky & Kahneman, *supra* note 65, at 154; Sangero & Halpert, *supra* note 59, at 50.

67. Sangero & Halpert, *supra* note 59, at 48.

68. Thomas Bayes, *An Essay Towards Solving a Problem in the Doctrine of Chances*, 53 PHIL. TRANSACTIONS ROYAL SOC’Y LONDON 370 (1763); Sangero & Halpert, *supra* note 59, at 49–50.

69. Sangero & Halpert, *supra* note 59, at 49; *see also, e.g.*, Spitalnic, *supra* note 43, at 55.

70. SHARON B. MCGRAYNE, THE THEORY THAT WOULD NOT DIE: HOW BAYES’ RULE CRACKED THE ENIGMA CODE, HUNTED DOWN RUSSIAN SUBMARINES, AND EMERGED TRIUMPHANT FROM TWO CENTURIES OF CONTROVERSY, at xi (2011).

evidence. And the third is the likelihood ratio. Simply defined, the likelihood ratio of a given body of evidence with respect to a given proposition is the ratio of the probability that the evidence would arise given that the proposition is true to the probability that the evidence would arise given that the proposition is false.⁷¹

If we return to John in our example, the accuracy of the test (99.9%) is a component of the likelihood ratio. The likelihood ratio here is the quotient of two conditional probabilities; the numerator is the probability that the test result will be positive given that the person tested is a carrier, and the denominator is the probability that the test result will be positive given that the person tested is not a carrier. If we assume a zero-probability of a false negative (i.e., a negative test result for a carrier), the following likelihood ratio results:⁷²

$$\text{Likelihood Ratio} = 1 / 0.001 = 1000$$

The Prior Odds here are the probability that a person is a carrier of the disease divided by the probability that he is not, *without* taking the test result into account. The Prior Odds for John, who is a member of a low-risk group with a base rate of 1-in-10,000, are as follows:

$$\text{Prior Odds} = 0.0001 / (1 - 0.0001) \cong 0.0001^{73}$$

The Posterior Odds are the probability that a person is a carrier divided by the probability that he is not, given a positive test result. If

71. Richard D. Friedman, *A Presumption of Innocence, Not of Even Odds*, 52 STAN. L. REV. 873, 875 (2000).

72. Sangero & Halpert, *supra* note 59, at 49.

73. *Id.* The denominator is supposed to complete the numerator to one, because the probability that a person is a carrier and the probability that he is not a carrier are complementary probabilistic occurrences. *Id.* at 49 n.33.

we now insert the figures we arrived at above into Bayes' Theorem, the resultant odds follow:

$$\text{Posterior Odds} = \text{Likelihood Ratio} \times \text{Prior Odds} = 1000 \times 0.0001 = 0.1$$

John's Posterior Odds, then, are 1-in-10: only one in eleven people who test positive from John's low-risk group will actually be a carrier, while ten of the eleven positive test results will be false. This is the same result we arrived at above without using Bayes' Theorem directly but by applying its underlying rationale.⁷⁴

As noted, the mistaken intuition that if the home HIV test is 99.9% accurate, that there is only a 0.1% probability that John is not a carrier if he tests positive (as opposed to the actual 91% probability of error), is the result of the base rate fallacy, which is also referred to as "the fallacy of the transposed conditional": rather than calculating the probability that a person is not a carrier given a positive test result (10/11 = ~91%), the probability of a non-carrier getting a positive test result is calculated (1/1,000 = 0.1%). From a Bayesian perspective, the source of this fallacy is that the Prior Odds are ignored and, consequently, the Posterior Odds are equated with the Likelihood Ratio. Daniel Kahneman and Amos Tversky have put this failure most succinctly: "The failure to appreciate the relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction."⁷⁵

3. *Applying Bayes' Theorem in the Criminal Justice System*

Lawyers, judges, and jurists are no less susceptible to these cognitive fallacies, for when they are relying on a single piece of evidence to convict a defendant, they are ignoring the Prior Odds and

74. *Id.* at 49–50.

75. Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, 80 *PSYCHOL. REV.* 237, 243 (1973); Sangero & Halpert, *supra* note 59, at 50.

the important distinction between conditional probability and inverse conditional probability.

In the criminal justice context, the given occurrence is a positive result yielded by scientific or other evidence. The probability of a person being a disease carrier in the medical diagnostic context is replaced by the probability of someone being guilty of a crime in the legal context; the probability of a person not being a carrier is replaced by the probability of a person being innocent. In medical diagnostics, the prior probability can be derived from the base rate; in criminal justice, the prior probability is the judge's (or the fact finder's) assessment of the defendant's guilt or innocence based not on the main evidence but on other evidence.⁷⁶

It is important to note here the debate as to whether Bayes' Theorem can and should be applied in criminal law, which began with the canonical exchange between Michael O. Finkelstein, William B. Fairley, and Laurence H. Tribe.⁷⁷ I am not, however, proposing that judges (or fact finders) apply Bayes' Theorem to make precise calculations to determine the cumulative weight of evidence. What I am suggesting, rather, is that the underlying logic of the formula be used to understand the risks of convicting a defendant based on one piece of evidence alone (of any type) and to persuade legislators to enact a safety rule that prohibits conviction based on a single piece of evidence.

To illustrate, consider two hypotheses and one given event.⁷⁸ Under the first hypothesis, the suspect's Guilt (designated "G") is assumed, and under the second hypothesis, the suspect's Innocence (designated "I") is assumed. The given event is the specific piece of Evidence incriminating the suspect (designated "E"), which could be

76. Sangero & Halpert, *supra* note 59, at 51.

77. See generally Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489, 489 (1970); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1329 (1971); see also Alex Stein, *Judicial Fact-Finding and the Bayesian Method: The Case for Deeper Skepticism About Their Combination*, 1 INT'L J. EVIDENCE & PROOF 25, 25 (1996); Sangero & Halpert, *supra* note 59, at 52; Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254, 1254 (2013).

78. Sangero & Halpert, *supra* note 34, at 539–41.

an eyewitness lineup identification or incriminating forensic lab test results. The basic odds formula is as follows:⁷⁹

$$\text{Likelihood Ratio} \times \text{Prior Odds} = \text{Posterior Odds}$$

The Likelihood Ratio (also known as the Bayes' Factor, which is used to update our prior beliefs with the evidence that we observe)⁸⁰ is the quotient of two conditional probabilities. The numerator is the probability of the existence of the evidence assuming the suspect is guilty, and the denominator is the probability of the existence of the evidence assuming the suspect is innocent.⁸¹ This is the mathematical expression of the strength of the evidence. For example, if the incriminating evidence is an eyewitness identification in a police lineup, then a Likelihood Ratio of ten means a likelihood the suspect was correctly identified as the guilty culprit ten times greater than the likelihood the suspect is innocent and mistakenly identified. But the Likelihood Ratio is not, by itself, a sufficient indicator of a suspect's guilt or innocence, as it does not take into account any other evidence aside from one specific piece of evidence, and instead presumes what is actually yet to be proven (the numerator assumes guilt and the denominator innocence).⁸²

The Prior Odds are the probability of a suspect's Guilt divided by the probability of his Innocence without taking into consideration the identification and the other admissible evidence.⁸³ These odds are called "Prior" because they reflect what we believe prior to observing the evidence.⁸⁴ Bayes' Theorem "updates" these prior beliefs by incorporating the evidence that we observe.⁸⁵ The Posterior Odds are then the product of the Likelihood Ratio times the Prior Odds.⁸⁶ This

79. *Id.* at 539–40.

80. Cheng, *supra* note 77, at 1267.

81. Sangero & Halpert, *supra* note 34, at 540.

82. *Id.*

83. *Id.*

84. Cheng, *supra* note 77, at 1266.

85. *Id.* at 1267.

86. *Id.*

represents the weight of the central piece of evidence combined with the other incriminating evidence against the suspect, expressed as follows:⁸⁷

$$\text{Posterior Odds} = P(\text{Guilt} \mid \text{Evidence}) / P(\text{Innocence} \mid \text{Evidence})$$

When the Posterior Odds are 1, the probability of guilt is identical to the probability of innocence (only a 50% likelihood that the suspect is guilty). The greater the Posterior Odds, the greater the probability of guilt.⁸⁸

Elsewhere, I have demonstrated with Dr. Mordechai Halpert the tremendous significance of not only the Likelihood Ratio but also the Prior Odds when a single piece of evidence is the basis for a conviction.⁸⁹ If we assume a Posterior Odds threshold of ninety as the minimum requirement for a conviction beyond a reasonable doubt, which is the criminal law standard of proof, then about 5% of the convictions will be false. The following table further demonstrates the significance of the Prior Odds and the Likelihood Ratio, using different numerical values in applying Bayes' Theorem:⁹⁰

87. Sangero & Halpert, *supra* note 34, at 540–41.

88. *Id.* at 541.

89. Sangero & Halpert, *supra* note 59, at 54–55.

90. *See* Sangero & Halpert, *supra* note 59, at 54 (modified version of the previous table).

Table 1: Probabilities of Guilt under Bayes' Theorem

| <i>Prior Odds</i> | \times <i>Likelihood Ratio</i> | = <i>Posterior Odds</i> | <i>Probability of Guilt Given the Evidence (rounded out)</i> |
|-----------------------|----------------------------------|-------------------------|--|
| 1 ("50:50") | 10 | 10:1 (10 G v. 1 I) | 91% |
| 1/10 (1 G v. 10 I) | 100 | 10:1 | 91% |
| 1/100 | 1,000 | 10:1 | 91% |
| 1/1,000 | 10,000 | 10:1 | 91% |
| 1/10,000 | 100,000 | 10:1 | 91% |
| 1/100,000 | 1,000,000 | 10:1 | 91% |
| 1/1,000,000 | 10,000,000 | 10:1 | 91% |
| 1/10,000 | 1,000 | 0.1:1 | 9% |

Table 1 demonstrates that even a slight possibility of error in a piece of evidence leads to the practical impossibility of showing guilt beyond a reasonable doubt based solely on that evidence (that is to say, in the absence of any other evidence that impacts the Prior Odds). For example, when the Prior Odds are 1-in-100,000 (or 1 in more than 100,000) only evidence that has an error rate below 1-in-1,000,000 will result in the desired Posterior Odds of 10 (probability of guilt 91%). Yet, in reality, no evidence has or can have such a low error rate and high level of accuracy. Moreover, the Prior Odds could be very low, one in millions, when there is exculpatory evidence such as an alibi. In such circumstances, for a conviction to be based upon this evidence, the accuracy of the (scientific) evidence should be an error rate of one error in several tens of millions of cases.

Finally, the bottom row of the table indicates the error rate that can be expected for realistic evidence, namely, 1-in-1,000. Thus, if the Prior Odds of guilt are 1-in-10,000, the Posterior Odds will be a mere 0.1:1. In other words, 91% of convictions based on scientific evidence with a 1:1000 error rate and Prior Odds of 1:10,000 will be false.

As long as a given piece of evidence has not been shown to meet a certain (unrealistic) very high precision requirement, it should not constitute the sole grounds for convicting the defendant in question.

4. *Insights and Lessons from the Field of Medical Diagnostics*

Table 2 sets out some fundamental definitions and formulas used in the medical diagnostic model:⁹¹

Table 2: The Generic 2 x 2 Table

| | <i>Has Condition</i> | <i>Does Not Have Condition</i> | |
|----------------------|--|---|---|
| <i>Test Positive</i> | A | B | Total Positive Tests (A + B) |
| <i>Test Negative</i> | C | D | Total Negative Tests (C + D) |
| | Number in Sample with Condition (A + C) | Number in Sample Without Condition (B + D) | Total Number of Subjects (A + B + C + D) |

91. With some minor modifications, the definitions, formulas, and table are from Spitalnic, *supra* note 43, at 54; see also Stuart Spitalnic, *Test Properties I: Sensitivity, Specificity and Predictive Values*, HOSP. PHYSICIAN, Sept. 2004, at 27, 27.

- “Sensitivity” refers to the probability that a test will be positive for a patient who has the tested-for condition (e.g., HIV).

$$\text{Sensitivity} = A / (A+C)$$

- “Specificity” refers to the probability that a test will be negative for a patient without the tested-for condition.

$$\text{Specificity} = D / (D+B)$$

- “Positive Predictive Value” (PPV) refers to the probability that a patient has the condition given a positive test result.

$$\text{PPV} = A / (A+B)$$

- “Negative Predictive Value” (NPV) refers to the probability that a patient does not have the condition given a negative test result.

$$\text{NPV} = D / (C+D)$$

Professionals in the testing diagnostics and measuring device manufacturing fields commonly use these above concepts in describing a test’s results.⁹² Because of the crucial importance of the data collected based on these concepts, FDA approval requires that manufacturers of medical devices not only provide data on a test’s Specificity and Sensitivity, but also on the Positive and Negative Predictive Values for the various populations,⁹³ and the information must be made public.⁹⁴ This is a quintessential safety measure, for, as demonstrated above in the HIV-test example, a test may be suitable

92. FOOD & DRUG ADMIN., *supra* note 44, at 7–8.

93. Sangero & Halpert, *supra* note 4, at 1298.

94. 21 U.S.C. § 360j(h)(4)(B) (2012).

for diagnosing high-risk populations but not (on its own at least) for diagnosing low-risk groups.

It is vital that those who will engage in safety in the legal field apply the error-prevention model developed and refined in the medical diagnostics field. The need for this is self-evident, first and foremost with regard to forensic evidence, which should not, in essence, be any different from all other scientific tests. The medical diagnostics model is suitable for weighing nonscientific evidence as well. Had even the most basic medical diagnostics model been adopted in the criminal justice system, it would have long arrived at the imperative safety rule that a conviction cannot be based on a single piece of evidence of any type. None of the types of evidence currently accepted in criminal law is suited to this function. The law's disregard for this problem is one of the central causes of the phenomenon of false convictions.

D. DNA Testing

As described in the 2009 NAS Report, “[u]nlike many forensic techniques that were developed empirically within the forensic [science] community, with little foundation in scientific theory or analysis, DNA analysis is a fortuitous byproduct of cutting-edge science.”⁹⁵ And certainly, DNA has strong probative value. But not even this type of evidence, considered the gold standard of forensic science, is error-free.

1. Random Match Probability

The basic assumption in DNA testing is that with the exception of identical twins, every person's DNA is unique. However, for the purposes of DNA forensic evidence, not all the molecules in the two DNA samples are tested and compared.

Andrea Roth has provided a clear and detailed description of the DNA testing process: “During forensic testing, DNA is amplified and

95. NAS-2009 Report, *supra* note 3, at 99.

typed at several locations, or loci, along the genetic strand.”⁹⁶ A DNA profile consists of two genetic markers (alleles) at each locus, representing the two alleles a person inherits from each of his two parents at that locus.⁹⁷ The main iteration of DNA-matching technology is called Short Tandem Repeat (STR).⁹⁸ The Federal Bureau of Investigation (FBI) and state laboratories use STR to test thirteen loci.⁹⁹ The DNA analysts use a statistical table developed by the FBI based on sample groups of approximately two hundred people from each of four racial categories to estimate the chances of finding each particular allele at each particular locus in the different racial groups.¹⁰⁰ Based on the assumption that the allelic frequencies among the loci are independent, the analyst multiplies the 26 (13x2) frequencies together to report for each group a “Random Match Probability” (RMP), the probability that a random person will have the twenty-six-allele profile.¹⁰¹

It is important to recall that the court is not called upon to determine the RMP, but rather the inverse conditional probability: namely, the probability that the two samples belong to the defendant and that, given a match, he is indeed guilty of the crime for which he is accused.¹⁰² To determine this probability, all the non-DNA-related evidence must be examined, either by applying the Bayes’ Theorem rationale or any other accepted method.¹⁰³ The erroneous belief that

96. Andrea Roth, *Safety in Numbers? Deciding When DNA Alone Is Enough to Convict*, 85 N.Y.U. L. Rev. 1130, 1135 (2010). For additional explanations of the scientific aspect of DNA testing, see David H. Kaye, *DNA Evidence: Probability, Population Genetics, and the Courts*, 7 HARV. J.L. & TECH. 101, 104 (1993); FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 129–210 (3d. ed. 2011); NAT’L COMM’N ON THE FUTURE OF DNA EVIDENCE, WHAT EVERY LAW ENFORCEMENT OFFICER SHOULD KNOW ABOUT DNA EVIDENCE, <https://www.ncjrs.gov/pdffiles1/nij/bc000614.pdf> [<https://perma.cc/E55R-QUJR>] (last visited July 7, 2018); Sangero & Halpert, *supra* note 59, at 72.

97. Roth, *supra* note 96, at 1135.

98. *Id.* at 1135–36.

99. *Id.* at 1136.

100. *Id.*

101. *Id.*

102. Sangero & Halpert, *supra* note 59, at 72.

103. *Id.*; see also Roth, *supra* note 96, at 1156.

the random match probability represents the probability of the defendant's innocence is termed the "prosecutor's fallacy."¹⁰⁴

The possibility of a random match in a DNA comparison is not particularly problematic in the context of a criminal trial, because it is well known that this probability must be taken into account.¹⁰⁵ Yet because the chances of a random match are at times expressed in astronomical terms (say, a one-in-a-billion probability of a match between the defendant's DNA and the crime scene DNA sample), courts run a considerably greater risk of falsely convicting an innocent defendant based on a DNA comparison when they ignore the possibility of lab error (as, unfortunately, is often the case).¹⁰⁶

The danger of error is far greater when the police have no specific suspect whose DNA profile they can compare to the sample from the crime scene (Verification), and instead must look for a match to one of the (millions) of DNA profiles in the DNA database (Identification or Database Search). This "cold hit" method can be expected to lead to false identifications,¹⁰⁷ and thus the results of such a database search must not be admissible as evidence but, rather, serve only as a means for finding suspects. Once a suspect is found, other significant evidence connecting him to the crime should be searched for; in the absence of such evidence, the suspect should be acquitted if brought to trial.¹⁰⁸ An important question, beyond the scope of this article, is what boundaries should be imposed on such database searches under Fourth Amendment protections.¹⁰⁹

104. Koehler, *supra* note 39, at 521.

105. Sangero & Halpert, *supra* note 59, at 73.

106. *People v. Reeves*, 109 Cal. Rptr. 2d 728, 753 (2001) ("[B]ecause appellant has not presented persuasive evidence of an ongoing controversy in the scientific community, we conclude that the NRC's recommendation is generally accepted, and DNA probability calculations need not be modified to account for a laboratory error rate."); Sangero & Halpert, *supra* note 59, at 79.

107. *But see* Rick Visser, *When DNA Won't Work*, 49 IDAHO L. REV. 39, 41 (2012).

108. Sangero & Halpert, *supra* note 59, at 53; *see also* Roth, *supra* note 96, at 1134. The author suggests a threshold of a 99.9% source probability (1-in-1,000) as satisfying the "beyond a reasonable doubt" requirement. *Id.* at 1168. I believe that this is not sufficient, as it relates only to the possibility of a random match and almost completely neglects the much greater possibility of a laboratory error.

109. *Maryland v. King*, 569 U.S. 435, 465–66 (2013); David H. Kaye, *A Fourth Amendment Theory of Arrestee DNA and Other Biometric Databases*, 15 U. PA. J. CONST. L. 1095, 1096 (2013); David H. Kaye, *Why So Contrived? Fourth Amendment Balancing, Per Se Rules, and DNA Databases after*

The use of partial DNA profiles is another problem in this context. Although the current U.S. standard is to test thirteen loci to arrive at a twenty-six-allele profile, partial DNA profiles containing fewer alleles can also be searched against profiles in databases¹¹⁰ (such as the FBI's Combined DNA Index System).¹¹¹ Using such partial DNA evidence to support prosecuting someone the police have identified as a suspect, particularly when other evidence already points to his guilt, is, in itself, not objectionable. Problems arise, however, when police reverse the order of their investigative methodology: when investigators screen a partial genetic profile against DNA profiles data to “trawl” for potential suspects, rather than matching the partial DNA profile taken from the crime scene against an already-identified suspect.¹¹² As Michael Naughton and Gabe Tan explain, three hazards arise.¹¹³ First, there is no way to estimate with sufficient certainty a match between the suspect's profile and the crime -scene profile if the latter is not complete.¹¹⁴ The authors illustrate this with the example of Raymond Easton, whose six STR markers matched the crime scene profile, but later on, when four other loci were tested, none was a fit.¹¹⁵ Second, the match probability increases significantly with a partial profile, and thus, speculative searches of partial crime scene profiles against DNA profiles in databases often produce multiple “matches.”¹¹⁶ Third, this problem makes innocent individuals whose DNA profiles are in the database more vulnerable to becoming suspects for a crime they did not commit and, accordingly, being falsely convicted.¹¹⁷

Maryland v. King, 104 CRIM. L. & CRIMINOLOGY 535, 535 (2014); Catherine W. Kimel, *DNA Profiles, Computer Searches, and the Fourth Amendment*, 62 DUKE L.J. 993, 993 (2013); Erin Murphy, *License, Registration, Cheek Swab: DNA Testing and the Divided Court*, 127 HARV. L. REV. 161, 161 (2013).

110. Michael Naughton & Gabe Tan, *The Need for Caution in the Use of DNA Evidence to Avoid Convicting the Innocent*, 15 INT'L J. EVIDENCE & PROOF 245, 251–53 (2011).

111. See NAS-2009 Report, *supra* note 3, at 100.

112. Naughton & Tan, *supra* note 110, at 252.

113. *Id.*

114. *Id.*

115. *Id.* at 252–53.

116. *Id.* at 253.

117. *Id.*

In an Arizona forensics laboratory in 2001, analysts found a nine-STR locus match between two unrelated individuals.¹¹⁸ The random match probability for a match of a nine-locus genotype in Arizona was “1 in 754 million in Caucasians” and “1 in 561 billion in African Americans.”¹¹⁹ In 2005, during proceedings in an Arizona court in a case in which the state had only typed nine loci, a DNA analyst testifying for the defense stated that she had found approximately *ninety* nine-locus, partial matches in a relatively small database with only 65,493 entries.¹²⁰ These findings seem to contradict the usual court testimony of forensic experts, who tend to estimate the RMP as one in millions, billions, or trillions.¹²¹ Keith Devlin, a mathematician at Stanford University, dismisses the extreme RMP numbers presented in courts as “nonsense” and “lies.”¹²² Moreover, some defense lawyers and researchers requested access to convicted-offender databases (such as the FBI’s CODIS and the National DNA Index System, or NDIS) to empirically test the theoretical estimates.¹²³ In response, the FBI threatened that if states opened their databases to external scientists or defendants, the FBI would terminate their participation in the national database system.¹²⁴ In his article, *Trawling DNA Databases for Partial Matches: What Is the FBI Afraid Of?*, David Kaye observed as follows:

[T]he release of the data, stripped of personal identifiers, for population-genetics research is permissible . . . the FBI has nothing to fear and should reverse its policy of not researching the issue and maintaining the secrecy of the data . . . the public and the legal community need to know that all reasonable efforts have been made to verify the

118. David H. Kaye, *Trawling DNA Databases for Partial Matches: What Is the FBI Afraid Of?*, 19 CORNELL J. L. & PUB. POL’Y 145, 153 (2009).

119. *Id.* at 154.

120. *Id.*

121. *See id.* at 155.

122. *Id.* at 148.

123. *Id.* at 149.

124. Kaye, *supra* note 118, at 149.

accuracy of the numbers that are given to police, judges, and juries. Disclosure of the databases in anonymized form is the best policy.¹²⁵

I agree with Kaye. A few researchers, including Bruce Weir and Laurence Mueller, have used simulations with databases in their research.¹²⁶ But the databases available to these researchers are relatively small.¹²⁷ I contend that conducting expanded simulations on the NDIS would be an important safety tool for the criminal justice system. Indeed, people should not be judged and sentenced to jail based on theories and RMP calculations alone when we can verify (using strong computers) the exact RMP for each number of loci in a profile.

Last, another significant problem is that many crime scene DNA samples contain a mixture of DNA from two or more people. This could also lead to the false identification of an innocent suspect as the perpetrator of the crime being investigated.¹²⁸ Indeed, the title of Naughton and Tan's article frames this concern in the very terms that underlie the objective of this article: "The Need for Caution in the Use of DNA Evidence to Avoid Convicting the Innocent."¹²⁹

A recent special report to the President discussed this problem of a mixture of DNA from two or more people.¹³⁰ The report distinguishes sharply between DNA analysis of single-source and simple-mixture samples and DNA analysis of complex-mixture samples.¹³¹ In the first category, DNA analysis "is an objective method in which the laboratory protocols are precisely defined and the interpretation involves little or no human judgement" (although

125. *Id.* at 150.

126. *Id.* at 161.

127. *Id.* at 155.

128. Naughton & Tan, *supra* note 110, at 254–55.

129. *Id.* at 245.

130. EXEC. OFFICE OF THE PRESIDENT, PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., REPORT TO THE PRESIDENT, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 7 (2016).

131. *Id.* at 7–8.

“errors can and do occur . . . sample mix-ups, contamination, incorrect interpretation, and errors in reporting”), whereas “*subjective analysis of complex DNA mixtures has not been established to be foundationally valid and is not a reliable methodology.*”¹³²

2. *Laboratory Error*

There is a significant risk of false conviction with DNA evidence due to the strong tendency to ignore the huge impact of a possible lab error in the DNA testing. The NAS 2009 report corrected this dangerous omission,¹³³ but as I will demonstrate, court verdicts have yet to reflect this.

The relatively recent use of genetic comparisons as scientific evidence in criminal trials did not emerge until the 1980s.¹³⁴ When attorneys first began to submit evidence from DNA testing in court, forensic experts claimed absolute reliability and zero probability of error in the tests.¹³⁵ Yet Koehler has rightly called for a distinction between a DNA match and a report of a DNA match based on their differing probabilities of error.¹³⁶ Indeed, the National Academy of Sciences found that:

Although DNA laboratories are expected to conduct their examinations under stringent quality controlled environments, errors do occasionally occur. They usually involve situations in which interpretational ambiguities occur or in which samples were inappropriately processed and/or contaminated in the laboratory. Errors also can occur when there are limited amounts of DNA, which

132. *Id.* The report describes efforts to develop computer programs that apply various algorithms to interpret complex mixtures in an objective manner and recommends development of objective methods. *Id.* at 78.

133. See NAS-2009 Report, *supra* note 3, at 132.

134. Kaye, *supra* note 96, at 101; Sangero & Halpert, *supra* note 59, at 72.

135. Thompson et al., *supra* note 45, at 47–48.

136. Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 868–69 (1996); Sangero & Halpert, *supra* note 59, at 73.

limits the amount of test information and increases the chance of misinterpretation. Casework reviews of mtDNA analysis suggest a wide range in the quality of testing results that include contamination, inexperience in interpreting mixtures, and differences in how a test is conducted.¹³⁷

Research has shown that a wide variety of factors can account for errors in DNA testing.¹³⁸ For example, cross-contamination and sample mix-ups can be chronic occurrences at even the best DNA laboratories.¹³⁹ The hazards and risks increase with Polymerase Chain Reaction (PCR) typing methods,¹⁴⁰ which entail the duplication of a small amount of DNA to produce a larger amount sufficient for conducting the DNA test.¹⁴¹ Here, even minute contamination of the small sample is likely to be dangerously amplified into a significant contamination of the enlarged sample, which biases the test results. Errors can occur at any phase of the testing, beginning with the sample-collecting stage and through to the actual test itself.¹⁴² In addition, the test involves subjective interpretation of lines that appear at its conclusion,¹⁴³ and an incorrect interpretation is likely to yield an erroneous result.¹⁴⁴ Even

137. NAS-2009 Report, *supra* note 3, at 132.

138. Sangero & Halpert, *supra* note 59, at 73.

139. Thompson, *supra* note 48, at 11.

140. Naughton & Tan, *supra* note 110, at 247; NRC-II Report, *supra* note 46, at 83–84.

141. The definition of PCR given in the FBI Standards for DNA Quality Assurance Standards for Forensic DNA Testing Laboratories (2009) is as follows:

Polymerase Chain Reaction (PCR) is an enzymatic process by which a specific region of DNA is replicated during repetitive cycles which consist of the following: (1) denaturation of the template; (2) annealing of primers to complementary sequences at an empirically determined temperature; and (3) extension of the bound primers by a DNA polymerase.

FED. BUREAU OF INVESTIGATION, QUALITY ASSURANCE STANDARDS FOR FORENSIC DNA TESTING LABORATORIES 6 (2011), <https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view> [<https://perma.cc/6QKJ-ZHHU>] [hereinafter FBI Standards for DNA].

142. NRC-II Report, *supra* note 46, at 87.

143. William C. Thompson et al., *Evaluating Forensic DNA Evidence: Essential Elements of a Competent Defense Review: Part I*, 27 CHAMPION 16, 18 (2003); NRC-II Report, *supra* note 46, at 84–85.

144. NRC-II Report, *supra* note 46, at 84–85.

the most human of errors, such as mislabeling samples, are possible in the best laboratories and even when the lab workers are certain that they have taken every precaution against error.¹⁴⁵

Some of the risks of scientific evidence are similar to those that arise with medical diagnostic tests,¹⁴⁶ but while the FDA fully regulates manufacturers of medical diagnostic devices, there is no such regulation of manufacturers of scientific evidence equipment, including DNA testing equipment, despite the many risks entailed.¹⁴⁷ For example, there is no error reporting duty for DNA testing equipment, which is accepted practice in safety-critical systems.¹⁴⁸ This lack of duty leads to unsafety.¹⁴⁹

An illustrative example is the user manual Applied Biosystems supplies with the DNA testing kits it manufactures,¹⁵⁰ which states as follows in bold lettering: “For Research, Forensic and Paternity Use Only. Not for use in diagnostic procedures.”¹⁵¹ What this means is that the same, lone piece of evidence that is a sufficient basis for convicting and sentencing someone to an extended prison term or even death is insufficient foundation for a medical diagnosis.¹⁵² Not

145. NAT’L RESEARCH COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE 89 (1992). For a deeper discussion of the causes of error, see Naughton & Tan, *supra* note 110, at 246–47; Thompson et al., *supra* note 45, at 1; James Herbie DiFonzo, *In Praise of Statutes of Limitations in Sex Offense Cases*, 41 HOUS. L. REV. 1205, 1233–61 (2004).

146. Sangero & Halpert, *supra* note 4, at 1305.

147. See François Pompanon et al., *Genotyping Errors: Causes, Consequences and Solutions*, 6 NATURE REV. GENETICS 847, 852–53 (2005); Thompson et al., *supra* note 45, at 10, 14. The FBI’s *Quality Assurance Standards for Forensic DNA Testing Laboratories* is important, but the only standard for DNA kits is imposed on laboratories and not manufacturers. FBI Standards for DNA, *supra* note 141, at 19 (“Standard 9.3[:] The laboratory shall identify critical reagents and evaluate them prior to use in casework. These critical reagents shall include but are not limited to the following: [] Test kits or systems for performing quantitative PCR and genetic typing.”).

148. Sangero & Halpert, *supra* note 4, at 1305.

149. *Id.*

150. *Id.* at 1306.

151. APPLIED BIOSYSTEMS, AMPFLSTR IDENTIFILER, PCR AMPLIFICATION KIT: USER’S MANUAL, at ii (2001), http://projects.nfstc.org/workshops/resources/literature/Ampflstr_Identifier_users_manual.pdf [<https://perma.cc/6PEQ-LD5M>]. The warning was recently slightly modified to “For Forensic or Paternity Use Only.” APPLIED BIOSYSTEMS, AMPFLSTR SGM PLUS PCR AMPLIFICATION KIT: USER’S MANUAL 2 (2012), http://www3.appliedbiosystems.com/cms/groups/applied_markets_support/documents/generaldocument/s/cms_041049.pdf [<https://perma.cc/E5N9-7Q6R>].

152. Sangero & Halpert, *supra* note 4, at 1306.

surprisingly, problems with Applied Biosystems software have emerged,¹⁵³ and it seems that the software was never approved by the FDA.¹⁵⁴

Accidents and incidents are not a rare phenomenon in DNA testing.¹⁵⁵ Exemplifying this is the widespread contamination discovered at the British Forensic Service. Researchers found that the DNA of twenty employees of the microfuge tubes manufacturer had contaminated the DNA evidence in scores of cases,¹⁵⁶ reporting that contamination had been found in approximately 10% of scenes.¹⁵⁷ A similar case arose in Germany. In 2008, the German police offered a 100,000 Euro award for information leading to the arrest of a serial killer known as the “Phantom of Heilbronn.”¹⁵⁸ Traces of her DNA had been found at some forty crime scenes in Germany, Austria, and France, six of them murders.¹⁵⁹ In 2009, it was revealed that in fact there had never been a serial killer, and instead, the DNA found at all the crime scenes belonged to an innocent female worker at the Bavarian factory that manufactures the cotton swabs used in the DNA collection.¹⁶⁰ The swabs had been contaminated with her DNA.¹⁶¹

Of course, optimists will maintain that the worst-case scenario of such cases of contamination is not the false accusation of an innocent

153. Jason R. Gilder et al., *Systematic Differences in Electropherogram Peak Heights Reported by Different Versions of the GeneScan Software*, 49 J. FORENSIC SCI. 92, 95 (2004); Sangero & Halpert, *supra* note 4, at 1306.

154. Sangero & Halpert, *supra* note 4, at 1306.

155. *Id.*

156. Kevin Sullivan et al., *New Developments and Challenges in the Use of the UK DNA Database: Addressing the Issue of Contaminated Consumables*, 146 FORENSIC SCI. INT’L 175, 176 (2004) (cited in Sangero & Halpert, *supra* note 4, at 1306).

157. *Id.*

158. Sangero & Halpert, *supra* note 4, at 1307; *Reward for “Phantom Killer” Reaches Record €300,000*, LOCAL (Jan. 13, 2009, 4:55 PM), <http://www.thelocal.de/national/20090113-16739.html> [<https://perma.cc/5NSN-M5LE>]; David H. Kaye, Commentary, *GINA’s Genotypes*, 108 MICH. L. REV. FIRST IMPRESSIONS 51, 52 (2010).

159. Kaye, *supra* note 158, at 52; Sangero & Halpert, *supra* note 4, at 1307.

160. Sangero & Halpert, *supra* note 4, at 1307.

161. *Reward for “Phantom Killer” Reaches Record €300,000*, *supra* note 158; Kaye, *supra* note 158, at 52; Sangero & Halpert, *supra* note 4, at 1307.

person, but that the actual perpetrator remains at large.¹⁶² But this is flawed thinking from the perspective of safety. First, safety in DNA testing could reduce the risk of false negatives, which are what allow the actual perpetrators to roam free. Second, safety in DNA testing would also reduce the risk of false positives that result from the relatively easy and undetectable sample contamination. A primary hazard is cross-contamination between the genetic matter of an innocent suspect (or someone who becomes a suspect after the DNA test) and the DNA sample taken from a crime scene, which leads to the mistaken conclusion that the suspect committed the crime.¹⁶³

In some cases, cross-contamination can lead to a false conviction, which almost occurred in the Jaidyn Leskie murder investigation¹⁶⁴ and the Russell John Gesah case,¹⁶⁵ and actually occurred in the Farah Jama case.¹⁶⁶ A prominent case in which cross-contamination almost led to a false conviction is that of Jack Bellamy, a convicted sex offender who was charged with the murder of Jane Durrua. In 2004, a DNA sample taken during the 1968 murder investigation was found to match Bellamy's DNA.¹⁶⁷ It later emerged, however, that

162. Sangero & Halpert, *supra* note 4, at 1307.

163. Thompson, *supra* note 48, at 10–12; Sangero & Halpert, *supra* note 4, at 1307.

164. In the Jaidyn Leskie murder investigation, DNA samples taken from a young, “mentally challenged” girl matched the samples taken from the murder scene. GRAEME JOHNSTONE, INQUEST INTO THE DEATH OF JAIDYN RAYMOND LESKIE, CORONERS CASE No. 007/98 64–65 (2006) (Austl.), http://darwin.bio.uci.edu/~mueller/pdf/leskie_decision.pdf [<https://perma.cc/C657-ABJM>]. It emerged in the coroner's inquiry that a sex crime committed against the girl had been investigated by the same laboratory that had tested the blood stains from the murder, which occurred around the same time. *Id.* at 67–70. The coroner's final conclusion was that there had been cross-contamination between the girl's DNA and the DNA sample from the Leskie murder scene. *Id.* at 70–72, 85. The coroner noted that additional instances of contamination had been discovered at the same laboratory. *Id.* at 85; Sangero & Halpert, *supra* note 4, at 1307–08.

165. In 2008, murder charges brought against Russell John Gesah based on DNA test results were dropped when it emerged that his DNA sample and the sample from the crime scene had been processed at the same time and by the same laboratory, raising cross-contamination concerns. Sangero & Halpert, *supra* note 4, at 1307–08.

166. Farah Jama was convicted and sentenced to six years in prison. FRANK H.R. VINCENT, REPORT: INQUIRY INTO THE CIRCUMSTANCES THAT LED TO THE CONVICTION OF MR. FARAH ABDULKADIR JAMA 13 (2010) (Austl.), <https://www.parliament.vic.gov.au/papers/govpub/VPARL2006-10No301.pdf> [<https://perma.cc/GYJ2-7KW6>]. Jama served about a year and a half until the prosecutor informed the court, in 2009, that the DNA sample had apparently been contaminated. *Id.* at 46. The court vacated the conviction, and Jama was released from prison. *Id.* at 47; Sangero & Halpert, *supra* note 4, at 1308.

167. Robert Hanley, *DNA Leads to Arrest in '68 Rape and Murder of Girl*, N.Y. TIMES (June 17,

the same laboratory processed Bellamy's original DNA sample and the sample from the Durrua investigation at the same time, giving serious reason to suspect cross-contamination,¹⁶⁸ and leading the prosecutor to drop the charges against Bellamy. Finally, in 2008, different results were achieved when the samples were tested by other laboratories, after which charges were filed against a new suspect, Robert Zarinsky, for the same murder.¹⁶⁹

Not all DNA testing errors are detected. In the case of John Ruelas and Gary Lieterman, for example, their DNA samples were found to match DNA found at a 1969 murder crime scene.¹⁷⁰ Ruelas, who had been four years old at the time of the murder, was clearly not the perpetrator.¹⁷¹ Lieterman, in contrast, was convicted of the murder, despite the lack of a reasonable explanation for the match between Ruelas' DNA and the DNA at the crime scene, and disregarding the fact that the sample from the victim and samples from the two suspects were processed in the same laboratory at the same time.¹⁷² Given the Hidden Accidents Principle in criminal law, had Ruelas been an adult and not a child at the time of the murder, he would likely have been falsely convicted and the probability of cross-contamination never revealed.¹⁷³

The media often reports on DNA testing mishaps, but does not always provide accurate information in doing so.¹⁷⁴ And as there is no reporting duty or duty to investigate DNA testing accidents—let alone incidents—the media is often the exclusive source of this information.¹⁷⁵ Consequently, again, not only does the criminal

2004), <http://www.nytimes.com/2004/06/17/nyregion/dna-leads-to-arrest-in-68-rape-and-murder-of-girl-13.html?sec=&spon=&pagewanted=all> [<https://perma.cc/2TUG-6YT2>]; Sangero & Halpert, *supra* note 4, at 1308.

168. William C. Thompson, *The Potential for Error in Forensic DNA Testing (and How That Complicates the Use of DNA Databases for Criminal Identification)*, COUNCIL FOR RESPONSIBLE GENETICS 28–29 (2008).

169. Sangero & Halpert, *supra* note 4, at 1308.

170. Thompson, *supra* note 48, at 14; Sangero & Halpert, *supra* note 4, at 1308–09.

171. Sangero & Halpert, *supra* note 4, at 1308.

172. *Id.* at 1309.

173. *Id.*

174. *Id.*

175. *Id.*

justice system lack a safety approach for preventing accidents, but it also makes no consistent attempt to learn from experience; that is, not even the outdated Fly-Fix-Fly method is applied in criminal law in a systematic fashion.¹⁷⁶

William C. Thompson has reported on the considerable errors and problems in how DNA laboratories are managed in the United States and elsewhere in both confirmation cases and “cold hit” database searches.¹⁷⁷ His findings are based on laboratory records and point to an unexpectedly high rate of detected cases of mislabeling and sample contamination.¹⁷⁸ Although the particular instances of laboratory contamination Thompson recorded were uncovered at early stages, he nonetheless noted that they raise grounds for concern, because cross-contamination is a regular occurrence even in the top-rated laboratories and “the same processes that cause detectable errors in some cases can cause undetectable errors in others.”¹⁷⁹ Thompson noted that “[e]rrors that incriminate a suspect are unlikely to be detected as errors; they are likely to be treated as incriminating evidence.”¹⁸⁰ He also considered the possibility of lab workers falsifying test results to cover up contamination incidents, which “can be the result of negligence, and cost a [lab] worker his job.”¹⁸¹

Koehler, in turn, has reported on professional proficiency tests that were not blind.¹⁸² “The error rates in these tests were tremendous, varying between 1% and 4%.”¹⁸³ Section II.C shows the tremendous significance of such an error rate. Moreover, Koehler, Audrey Chia, and Samuel Lindsey claimed that when the probability of a laboratory error is much greater than the RMP, the latter probability

176. *Id.*

177. Thompson, *supra* note 48, at 11.

178. *Id.* at 13.

179. *Id.* at 12.

180. *Id.*

181. *Id.*; see also Richard O. Lempert, *After the DNA Wars: A Mopping Up Operation*, 31 *ISR. L. REV.* 536, 552–53 (1997); Sangero & Halpert, *supra* note 59, at 75.

182. Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence*, 34 *JURIMETRICS J.* 21, 25–26 (1993); Sangero & Halpert, *supra* note 59, at 75–76.

183. Koehler, *supra* note 182, at 26; Sangero & Halpert, *supra* note 59, at 75.

is insignificant and the former probability is the relevant statistic.¹⁸⁴ To illustrate, an RMP of 1-in-100-million creates a bias against the defendant, even if the probability of a laboratory error is 1%, because the RMP is the only statistic the jury hears. It is therefore preferable not to report the RMP to the jury. In a later article, Koehler made the recommendation—which I fully support—that the jury instructions on the possibility of an error in the testing include only one statistic, relating to both the RMP and lab error combined.¹⁸⁵

Another claim Koehler raised is that the average error rate of all forensic laboratories should be considered in the absence of statistics regarding the error rate of the laboratory that performed the actual testing.¹⁸⁶ I take the further step, however, of maintaining that the absence of statistical data on the error rate of the relevant laboratory should render its test results inadmissible as criminal evidence given, among other things, the *Daubert* rule. Alternatively, if the court nonetheless admits the results as evidence, safety considerations mandate that the error rate of the relevant laboratory be assumed to be at the highest level for laboratories of the same type.¹⁸⁷

It has been claimed that retesting can reduce the error rate, particularly if it is performed by a different laboratory,¹⁸⁸ but this would not reduce the error rate to zero. First of all, the same cause of error in the first round of testing could quite possibly reoccur in the retesting at the second laboratory;¹⁸⁹ indeed, different laboratories have been known to make the same mistakes.¹⁹⁰ There are a variety of reasons for the same mistake to be repeated in different laboratories, including the erroneous analysis of the lines obtained in

184. Jonathan J. Koehler et al., *The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?*, 35 JURIMETRICS J. 201, 210–11 (1995); Sangero & Halpert, *supra* note 59, at 76.

185. Koehler, *supra* note 39, at 533.

186. Jonathan J. Koehler, *Why DNA Likelihood Ratios Should Account for Error (Even When a National Research Council Report Says They Should Not)*, 37 JURIMETRICS J. 425, 433 (1997).

187. Sangero & Halpert, *supra* note 59, at 76. A similar claim was raised in Barry C. Scheck, *DNA and Daubert*, 15 CARDOZO L. REV. 1959, 1981–85 (1994).

188. NRC-II Report, *supra* note 46, at 37; Sangero & Halpert, *supra* note 59, at 78.

189. David J. Balding, *Errors and Misunderstandings in the Second NRC Report*, 37 JURIMETRICS J. 469, 475 (1997).

190. Koehler, *supra* note 186, at 437; Thompson et al., *supra* note 45, at 2.

the test and contamination of the sample before the first test.¹⁹¹ Another problem is that the first laboratory sometimes uses up all of the sample material, making retesting impossible.¹⁹²

Despite this, retesting could still prevent certain laboratory errors that lead to false convictions. Thus, as a necessary safety procedure, a legal rule should be passed requiring retesting DNA samples by an independent, objective expert as a condition for basing a conviction on DNA testing lab results.¹⁹³ However, DNA evidence should not suffice alone for convictions, because performing a second test will not neutralize altogether the significant risk of error.

Finally, Israeli researchers exposed the possibility of DNA fabrication as another source of concern when they created artificial DNA that can fool current forensic testing procedures.¹⁹⁴ Following this, it was suggested that “the discovery of the ability to easily fabricate DNA evidence as well as a long history of DNA falsification and gross ineptness by crime laboratories demonstrate that DNA-based evidence’s sterling reputation is undeserved.”¹⁹⁵

3. *Summary and Recommendations*

Although it is undisputed that DNA evidence is significant, weighty evidence that the courts must rely on, it must not be allowed to constitute the sole basis for a conviction in a criminal trial, because sole reliance creates a tangible danger that the conviction will be wrongful.¹⁹⁶

Indeed, as discussed, the likelihood of a lab error in DNA testing tends to be considered in detachment from the other evidence in a

191. Thompson et al., *supra* note 45, at 48.

192. This was common practice in the Houston police crime laboratory. DiFonzo, *supra* note 145, at 1248; Sangero & Halpert, *supra* note 59, at 78.

193. Sangero & Halpert, *supra* note 59, at 21.

194. D. Frumkin et al., *Authentication of Forensic DNA Samples*, 2009 FORENSIC SCI. INT’L.: GENETICS 1, 1; Bolden, *supra* note 11, at 409.

195. Bolden, *supra* note 11, at 440–41 (“[C]ourts should evaluate DNA evidence on a case-by-case basis, evaluating the authenticity of the DNA evidence as well as the testing procedures used to obtain the results.”).

196. Sangero & Halpert, *supra* note 59, at 80.

case,¹⁹⁷ even though such errors are unavoidable and the court has no way of determining whether these errors occurred in the specific case at issue. Not even retesting in another laboratory will fully fix this problem. Moreover, although the probability of a lab error (which may occur in 1 in 100 cases) is much higher than the RMP, juries are not supplied with this statistic and instead hear only the impressive RMP statistic (which could amount to one-in-a-billion or even trillion cases).

Because there are no adequate statistics on the error rates of different laboratories, the prosecution should bear the burden of establishing these rates regarding the specific laboratory that performed the testing upon which the prosecution relies.¹⁹⁸ Above I suggested two alternative safety mechanisms when there is a lack of data on a specific laboratory; the court must either find the evidence inadmissible or admit it while ascribing the highest known rate of error for tests performed in laboratories of the relevant sort.

In 2009, the FBI released its *Quality Assurance Standards for Forensic DNA Testing Laboratories*.¹⁹⁹ These groundbreaking standards, based partially on the recommendations in the NAS reports,²⁰⁰ include requirements for protocols regarding, among other things, the interpretation of DNA, mixed samples, and contamination.²⁰¹ Also included are important standards relating to quality assurance programs, education and training of laboratory personnel, lab reports, lab reviews, proficiency testing, and corrective

197. *Id.* at 79.

198. *Id.*

199. *Revised Quality Assurance Standards*, NAT'L INST. OF STANDARDS & TECH., <https://strbase.nist.gov/revisedQAS.htm> [<https://perma.cc/283A-3DEW>] (last visited July 7, 2018). For the *Quality Assurance Standards* currently in force, see FBI Standards for DNA, *supra* note 141.

200. See FBI Standards for DNA, *supra* note 141, at 20 (“Standard 9.6[:] The laboratory shall have and follow written guidelines for the interpretation of data. [] For a given population(s), the statistical interpretation of autosomal loci shall be made following the recommendations 4.1, 4.2 or 4.3 as deemed applicable of the National Research Council report entitled ‘The Evaluation of Forensic DNA Evidence’ (1996) and/or court directed method.”)

201. *Id.* at 20–21 (Standards 9.6 and 9.7).

action when discrepancies are detected in proficiency tests and casework analysis.²⁰²

Although this is important progress, these are only the first steps forward, as these standards should be mandatory for each and every federal and state laboratory. This has the potential to be an effective safety program, but to achieve this the standards must not simply require that labs formulate their own protocols, the standards must instead actually formulate in detail—and enforce—the necessary protocols.

In addition, a regulatory regime similar to the mandatory premarket approval process for medical diagnostic devices should be instituted for manufacturers of scientific evidence devices, including DNA kits.²⁰³ This regime should impose an accident-reporting duty as well as a duty to report incidents that involve accuracy, similar to the arrangement for medical devices. This would supplement and support existing safety recommendations relating to accreditation of laboratories, as set forth in the 2009 NAS Report.²⁰⁴

Another important way to ensure safety in the context of DNA testing and evidence is the performance of extended simulations on the NDIS. Using strong computers, researchers should verify and find out the exact RMP for each number of loci in a profile. Finally, the above-mentioned recommendation of the Report to the President (2016) should be adopted: to develop objective methods of interpreting complex DNA mixtures.²⁰⁵

E. Fingerprint Comparisons

1. The Possibility of Error

There are four stages to the basic approach of latent fingerprint experts, known as ACE-V: Analysis, Comparison, Evaluation, and

202. *Id.* at 10, 22–26 (Standards 5, 11–14).

203. This recommendation was originally made in Sangero & Halpert, *supra* note 4, at 1322.

204. See NAS-2009 Report, *supra* note 3, at 195.

205. EXEC. OFFICE OF THE PRESIDENT, *supra* note 130, at 15.

Verification.²⁰⁶ In the analysis stage, the examiner closely examines the latent print associated with the crime being investigated and decides whether there is enough useful information contained in the image to make it “of value” for further examination.²⁰⁷ If there is, the examiner marks up the print and documents the minutiae he or she observes.²⁰⁸ In the comparison stage, the analyst compares the latent print to a particular source print, noting observed similarities and differences.²⁰⁹ In the evaluation stage, the expert reaches one of three possible conclusions: exclusion, identification, or inconclusive.²¹⁰ If the first expert reaches an identification conclusion, then a second expert conducts the same process in the verification stage.²¹¹

Fingerprint evidence has long been considered very strong evidence.²¹² Throughout the twentieth century, both courts and the general public regarded it as the epitome of reliable and certain evidence, and it served as a basis for many convictions.²¹³ Yet in recent years, this special status has become the subject of criticism for not being grounded in solid statistical theory and for being subject to error.²¹⁴

As discussed above, general consensus exists as to the possibility of a random match in a DNA comparison so that all of the loci compared in a test will be identical for a number of people. For this

206. Mnookin, *supra* note 30, at 1217.

207. *Id.*

208. *Id.*

209. *Id.* at 1217–18.

210. *Id.* at 1218.

211. *Id.*

212. Mnookin, *supra* note 30, at 1217 (“Fingerprint evidence is, in all likelihood, both more probative and less error-prone than some other kinds of forensic identification evidence, and it has a long and extremely substantial courtroom use.”).

213. Robert Epstein, *Fingerprints Meet Daubert: The Myth of Fingerprint “Science” Is Revealed*, 75 S. CAL. L. REV. 605, 605 n.3 (2002); Sangero & Halpert, *supra* note 59, at 63.

214. Epstein, *supra* note 213, at 622; Zabell, *supra* note 17, at 152; Nathan Benedict, *Fingerprints and the Daubert Standard for Admission of Scientific Evidence: Why Fingerprints Fail and a Proposed Remedy*, 46 ARIZ. L. REV. 519, 526–33 (2004); Jennifer L. Mnookin, *Fingerprint Evidence in the Age of DNA Profiling*, 67 BROOK. L. REV. 13, 57–61 (2001); Sangero & Halpert, *supra* note 59, at 63–64; Mnookin, *supra* note 30, at 1209–10; Elizabeth J. Reese, *Techniques for Mitigating Cognitive Biases in Fingerprint Identification*, 59 UCLA L. REV. 1252, 1255–56 (2012); Brandon Garrett & Gregory Mitchell, *How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment*, 10 J. EMPIRICAL LEGAL STUD. 484, 485 (2013).

reason, the test results are given in statistical form: in a population of X million people, on average, Y persons will share the same genetic profile. However, the prevailing assumption regarding fingerprint comparisons is that every fingerprint is unique and there is zero possibility of a random match.²¹⁵ Consequently, courts tend not to require random match data for fingerprints and the prosecution therefore does not present any such data during trial. In effect, no data addresses this possibility,²¹⁶ and there is no scientific proof that it is impossible for two people to have the same points of comparison in a fingerprint examined by an expert.²¹⁷ Forensic experts testifying in court present this evidence as unequivocal instead of making an effort to investigate and provide data about Random Match Probability. This testimony leads jurors to perceive this evidence as far stronger than it actually is.²¹⁸

A 2002 study arrived at a 6.10×10^{-8} probability of a fingerprint with thirty-six minutiae points sharing twelve minutiae points with another arbitrarily chosen fingerprint with thirty-six minutiae points.²¹⁹ Thus, some statistical theories have found a possibility of a random match between fingerprints, similar to cases with DNA comparisons.²²⁰ Examiners comparing two different peoples' fingerprints may find them so similar that they cannot distinguish between them.²²¹

215. See generally David J. Balding & Peter Donnelly, *Inferring Identity from DNA Profile Evidence*, 92 PROC. NAT'L. ACAD. SCI. 11741 (1995); Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 MICH. L. REV. 931 (1999); see also Thompson et al., *supra* note 45, at 49–51; William C. Thompson & Simon A. Cole, *Psychological Aspects of Forensic Identification Evidence*, in EXPERT PSYCHOLOGICAL TESTIMONY FOR THE COURTS 31, 56 (Mark Costanzo et al. eds., 2007); Sangero & Halpert, *supra* note 59, at 64–65.

216. See Mnookin, *supra* note 30, at 1221–27; Zabell, *supra* note 17, at 155–56.

217. Mnookin, *supra* note 30, at 1225; Sharath Pankanti et al., *On the Individuality of Fingerprints*, 24 IEEE TRANSACTIONS ON PATTERN ANALYSIS & MACHINE INTELLIGENCE 1010, 1011 (2002).

218. Mnookin, *supra* note 30, at 1226; Saks & Koehler, *Forensic Identification Science*, *supra* note 35, at 893; Sangero & Halpert, *supra* note 59, at 64.

219. Pankanti et al., *supra* note 217, at 1021; Sangero & Halpert, *supra* note 59, at 64.

220. Cedric Neumann et al., *Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Three Minutiae*, 51 J. FORENSIC SCI. 1255, 1255 (2006).

221. This was proven in the Brandon Mayfield case, discussed *supra* Section II.E.(2). See Sarah Kershaw et al., *Spain and U.S. at Odds on Mistaken Terror Arrest*, N.Y. TIMES (June 5, 2004), <http://www.nytimes.com/2004/06/05/us/spain-and-us-at-odds-on-mistaken-terror-arrest.html>

In the early 1990s, British researchers examined the sixteen-point standard for comparing fingerprints followed in England and Wales.²²² Their research findings showed the subjective nature of fingerprint analysis: different examiners arrived at entirely different points and numbers of comparison.²²³ The results of proficiency tests for 156 fingerprint examiners, conducted in the United States under the auspices of the International Association for Identification and published in 1996, reinforced this outcome.²²⁴ These results shocked the forensic science community: of the 156 examiners tested, only sixty-eight had both correctly identified the five latent print impressions that they were supposed to identify and correctly noted the two elimination latent prints that they were not supposed to identify.²²⁵ In total, scientists counted forty-eight false matches.²²⁶ The combined results of these proficiency tests show that fingerprint examiners get erroneous results in an average of 0.8% of cases²²⁷—a significant error rate.

Of course, errors in fingerprint analysis also occur in actual cases before the courts. Simon Cole reviewed twenty-two documented cases in the United States, England, and Scotland in which people were arrested and, at times, even served prison sentences before the error was detected.²²⁸ Considering the Hidden Accidents Principle in criminal law, this is likely only the tip of the iceberg of errors in

[<https://perma.cc/FUN2-7D4U>]; Sangero & Halpert, *supra* note 59, at 65.

222. I.W. Evett & R.L. Williams, *A Review of the Sixteen Points Fingerprint Standard in England and Wales*, 46 J. FORENSIC IDENTIFICATION 49, 49 (1996); Sangero & Halpert, *supra* note 59, at 65.

223. Evett & Williams, *supra* note 222, at 72; Sangero & Halpert, *supra* note 59, at 65; *see also* Mnookin, *supra* note 30, at 1221–22.

224. David L. Grieve, *Possession of Truth*, 46 J. FORENSIC IDENTIFICATION 521, 523 (1996); Sangero & Halpert, *supra* note 59, at 65.

225. Grieve, *supra* note 224, at 524.

226. *Id.*

227. LYN HABER & RALPH N. HABER, *Error Rates for Human Fingerprint Examiners*, in AUTOMATIC FINGERPRINT RECOGNITION SYSTEMS 339, 349 (Nalini K. Ratha et al. eds., 2003); Cole, *supra* note 42, at 1034, 1073; *see* Joseph L. Peterson & Penelope N. Markham, *Crime Laboratory Proficiency Testing Results, 1978–1991, II: Resolving Questions of Common Origin*, 40 J. FORENSIC SCI. 1009 (1995); Sangero & Halpert, *supra* note 59, at 65.

228. Cole, *supra* note 42, at 1001–16; Sangero & Halpert, *supra* note 59, at 65–66.

fingerprint analysis, but most errors have not been detected and the falsely convicted inmates remain in prison.

A number of issues may cause frequent laboratory errors in fingerprint analysis, including: poor quality of fingerprints taken from the crime scene (as opposed to the good quality of prints calmly scanned by access control systems),²²⁹ automated fingerprint identification systems,²³⁰ substandard or unscientific practices among certain “experts,”²³¹ and pressure exerted on laboratory staff by the police and/or prosecution to find a match.²³² In addition, latent images are often distorted, smaller in surface area than the full print, and frequently contain artifacts resulting from the processes necessary to make latent prints visible. As a result, two impressions from two different sources could be mistaken as coming from the same source.²³³

Leading forensic science researchers have called for the abandonment of “absolute conclusions” and, instead urge for the recognition of the inherently probabilistic nature of fingerprint evidence.²³⁴ The key question is not the uniqueness of friction ridge skin, but rather the fingerprint examiner’s ability to derive sufficient information from very limited sources. The researchers have suggested replacing experience and tradition alone with transparent and empirically-based practice.²³⁵ Yet, as Mnookin describes the current situation,

ACE-V’s relationship to the scientific method is tenuous at

229. See Pankanti et al., *supra* note 217, at 1016; Sangero & Halpert, *supra* note 59, at 66.

230. See U.S. DEP’T OF JUSTICE, OFFICE OF THE INSPECTOR GEN., OVERSIGHT & REVIEW DIV., A REVIEW OF THE FBI’S HANDLING OF THE BRANDON MAYFIELD CASE 1 (2006), <http://www.justice.gov/oig/special/s0601/exec.pdf> [<https://perma.cc/6M2S-GW8V>] (unclassified and redacted).

231. Saks & Koehler, *Forensic Identification Science*, *supra* note 35, at 893.

232. *Id.*; Sangero & Halpert, *supra* note 59, at 66.

233. Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725, 751 (2011).

234. Christopher Champod & Ian W. Evett, *A Probabilistic Approach to Fingerprint Evidence*, 51 J. FORENSIC IDENTIFICATION 101, 110 (2001).

235. *Id.*; see also Mnookin et al., *supra* note 233, at 751.

best . . . [L]atent fingerprint examination as a field lacks any formalized specifications about what is required in order to declare a match. There is no required minimum number of points of resemblance or minimum number of total print features, nor any required quantum of any specific kind of ridge detail . . . Two fingerprint analysts will often focus on different minutiae in their examination of the same print . . . The judgment is fundamentally a subjective one, not based on any formalized measures of either quantity or sufficiency. Additionally, latent fingerprints examiners do not generally employ any statistical information or models in the ordinary ACE-V process . . . [T]here simply is no well-accepted, fully-specified statistical model that is available for latent fingerprint examiners to employ . . . [F]undamentally, fingerprint matching ought to be thought of as a probabilistic inquiry.²³⁶

2. *The Brandon Mayfield Case and the U.S. Department of Justice's Report*

Following the 2004 terror attacks in Madrid, which led to 191 deaths and 2,050 nonfatal injuries, the Spanish police found a fingerprint on a blue plastic bag near one of the attack sites; it contained detonators and explosives remnants.²³⁷ The Spanish police requested assistance from the FBI,²³⁸ which searched its fingerprint database using an Automated Fingerprint Identification System (AFIS). The search printout identified twenty potential suspects.²³⁹ In analyzing the samples from the possible suspects, a fingerprint examiner found a match between the print taken from the attack site and that of a Portland, Oregon, attorney named Brandon Mayfield.²⁴⁰

236. Mnookin, *supra* note 30, at 1219–22.

237. U.S. DEP'T OF JUSTICE, *supra* note 230, at 1; Sangero & Halpert, *supra* note 59, at 66–67.

238. U.S. DEP'T OF JUSTICE, *supra* note 230, at 1.

239. *Id.*

240. *Id.*

Mayfield told FBI interrogators that he had never been to Spain in his life, had been in the United States at the time of the attacks, and did not even have a passport.²⁴¹ Nonetheless, three senior FBI examiners verified the identification of Mayfield's fingerprints,²⁴² and the affidavit supporting an arrest warrant for Mayfield in the United States declared a 100% positive identification.²⁴³ While Mayfield was in detention, the court appointed an independent fingerprint examiner to verify the identification made by FBI examiners.²⁴⁴ Two weeks after Mayfield's arrest, the Spanish police located someone else, an Algerian named Ouhnane Daoud, whose fingerprints matched the prints found on the plastic bag at the scene of the attacks.²⁴⁵ Thus, the 100% "certain" identification of Mayfield's fingerprints by four different examiners was wrong. Mayfield was released²⁴⁶ and subsequently received \$2 million in compensation.²⁴⁷

In 2006, the U.S. Justice Department released a comprehensive report on the Mayfield case.²⁴⁸ The report stated that the main cause of the false identification was the very strong similarity between Mayfield's fingerprint and the print from the attack site, which, according to Spanish police, belonged to Ouhnane Daoud.²⁴⁹ The report explained that it is possible for a great similarity between fingerprints to arise in an AFIS search, as the system scans millions of prints and compares each to the prints found at the scene of the crime.²⁵⁰ The system produces a list of twenty candidates who it has found to have the most similar fingerprints to the crime scene

241. Mnookin, *supra* note 30, at 1228.

242. U.S. DEP'T OF JUSTICE, *supra* note 230, at 1–2.

243. *See id.* at 18–19.

244. *Id.* at 3; Sangero & Halpert, *supra* note 59, at 66–67.

245. U.S. DEP'T OF JUSTICE, *supra* note 230, at 3.

246. Mnookin, *supra* note 30, at 1228–30; Sangero & Halpert, *supra* note 59, at 66–67; Robert B. Stacey, *Report on Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 54 J. FORENSIC IDENTIFICATION 706, 706 (2004); William C. Thompson & Simon A. Cole, *Lessons from the Brandon Mayfield Case*, 29 CHAMPION 42, 42 (2005); Steven T. Wax & Christopher J. Schatz, *A Multitude of Errors: The Brandon Mayfield Case*, 28 CHAMPION 6, 10 (2004).

247. Mnookin, *supra* note 30, at 1229.

248. U.S. DEP'T JUSTICE, *supra* note 230, at 1.

249. *Id.* at 3, 6.

250. *Id.* at 1.

fingerprints.²⁵¹ Consequently, the fingerprint examiners had to analyze fingerprints that were very similar to those of Daoud. The report cautioned that the risk of error with “cold hit” database searches is far greater than when suspects are identified by way of a regular police investigation,²⁵² and the constantly growing size of the databases increases the risk of misidentification.²⁵³

Peer pressure and “expectation bias” also played a role in the misidentification; once the first expert has declared a match between prints, the verifying experts naturally expect to find the same match.²⁵⁴ And indeed, the Justice Department report determined that the second verifying examiner knew that the first examiner had found a match between Mayfield’s print and the fingerprints from the scene of the crime.²⁵⁵ Thus, the report recommended withholding such information from verifying examiners.²⁵⁶

Another factor in the match found in the misidentification of Mayfield’s prints is cognitive bias. Mayfield had converted to Islam sometime earlier, his wife was Egyptian, and he had once represented a known terrorist in a child custody dispute.²⁵⁷ The report determined that this background information did not influence the initial identification of a match, as it was unknown to police investigators at the time, but it did impact the verification stage when examiners had this data.²⁵⁸ Forensic examiners often have access to or are provided with external information about the case that is irrelevant to the analysis,²⁵⁹ and research has shown that this impacts the analysis. Itiel E. Dror et al. used the Mayfield case to prove “contextual bias” in these circumstances. Five experts were each given a different pair of latent prints and potential source prints and told that they were the

251. *Id.*

252. *Id.* at 7.

253. Mnookin, *supra* note 30, at 1229.

254. *Id.* at 1230.

255. U.S. DEP’T JUSTICE, *supra* note 230, 10–11.

256. *Id.* at 11.

257. *Id.* at 2.

258. *Id.* at 11–12.

259. Mnookin, *supra* note 30, at 1230.

prints from the well-known Mayfield case, when in fact each expert received a pair of prints that the same expert had previously analyzed, identified, and testified to in court as a 100% match.²⁶⁰ In the Dror et al. experiment, three of the experts arrived at the opposite conclusion of no match, likely due to bias; one expert found the prints “inconclusive”; and only one was consistent and again identified a match between the prints.²⁶¹ Other experiments conducted by Dror et al. with different experts also found bias.²⁶²

Although the medical diagnostics and other scientific fields make formalized efforts to shield researchers from “contextual information,” no such procedures generally exist in the forensic sciences field.²⁶³ Thus, to prevent some of the biases, I suggest giving fingerprint examiners not just the suspect’s print to compare against the latent print from the crime scene, but also several “filler” prints from other people as well. This will prevent the examiner from knowing at the outset which print belongs to the suspect.²⁶⁴

As we then see, errors in fingerprint analysis can and do (as in the Mayfield case) actually occur in reality and are not a mere theoretical probability. Moreover, given the Hidden Accidents Principle, there are likely many more cases in which these errors remain undetected.²⁶⁵ Compounding this problem is the fact that some fingerprints are so similar to one another that examiners are incapable of distinguishing between them. Thus, the general error rate in fingerprint analysis is unknown.²⁶⁶ Although the general error rate is

260. Itiel E. Dror, David Charlton & Ailsa E. Péron, *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 *FORENSIC SCI. INT’L* 74, 75 (2006).

261. *Id.* at 76; see also Mnookin, *supra* note 30, at 1232.

262. Itiel E. Dror & D. Charlton, *Why Experts Make Errors*, 56 *J. FORENSIC IDENTIFICATION* 600, 600 (2006); Itiel E. Dror & Robert Rosenthal, *Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts*, 53 *J. FORENSIC SCI.* 900, 903 (2008); see also Mnookin, *supra* note 30, at 1232.

263. Mnookin, *supra* note 30, at 1230–31.

264. See Gary L. Wells et al., *Forensic Science Testing: The Forensic Filler-Control Method for Controlling Contextual Bias, Estimating Error Rates, and Calibrating Analysts’ Reports*, 2 *J. APPLIED RES. MEMORY & COGNITION* 53, 53 (2013) (offering a similar recommendation); see also William C. Thompson, *What Role Should Investigative Facts Play in the Evaluation of Scientific Evidence*, 43 *AUSTRALIAN J. FORENSIC SCI.* 123, 123 (2011).

265. Sangero & Halpert, *supra* note 59, at 68.

266. Mnookin, *supra* note 30, at 1227–28; Mnookin, *supra* note 214, at 59.

commonly estimated to be low, proficiency tests given to examiners belie this estimate. Above in Section II.C we have seen the tremendous significance of even a very low error rate.

3. Case Law

Fingerprint analysis evidence is generally sufficient as the sole basis for a conviction in American courts.²⁶⁷ Over the last decade and a half, courts have begun to indicate skepticism with regard to this type of evidence,²⁶⁸ with some explicitly holding errors to be possible in fingerprint comparisons.²⁶⁹ However, the problematic report referred to by the FBI as the “50K Study” gave undeserving support to fingerprint evidence.²⁷⁰ Although the findings of this so-called study were never published in a scientific journal or subject to peer review (as the *Daubert* standard requires),²⁷¹ judges have nonetheless relied on these findings in their rulings.²⁷²

The “study” was in fact conducted to find support for the claim relied on by prosecutors that every fingerprint is unique and that false positive errors are not possible in fingerprint comparisons.²⁷³ FBI examiners used an AFIS search with a computerized database of 50,000 fingerprints to compare each fingerprint against itself and against the 49,999 other fingerprints in the database.²⁷⁴ This process

267. See, e.g., *People v. Rhodes*, 422 N.E.2d 605, 608 (Ill. 1981); *People v. Ford*, 606 N.E.2d 690, 693 (Ill. App. Ct. 1992); see also Sangero & Halpert, *supra* note 59, at 69.

268. See *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 565 (E.D. Pa. 2002); see also Epstein, *supra* note 213, at 269; Simon A. Cole, *Grandfathering Evidence: Fingerprint Admissibility Rulings from Jennings to Llera Plaza and Back Again*, 41 AM. CRIM. L. REV. 1189, 1196–97 (2004); David H. Kaye, *The Nonscience of Fingerprinting: United States v. Llera-Plaza*, 21 QUINNIPIAC L. REV. 1073, 1074 (2003); Michael Lynch & Simon A. Cole, *Science and Technology Studies on Trial: Dilemmas of Expertise*, 35 SOC. STUD. SCI. 269, 270 (2005); Mnookin, *supra* note 22, at 48; Sangero & Halpert, *supra* note 59, at 69.

269. *United States v. Crisp*, 324 F.3d 261, 273–74 (4th Cir. 2003) (Michael, J., dissenting); *State v. Quintana*, 103 P.3d 168, 171 (Utah Ct. App. 2004); Sangero & Halpert, *supra* note 59, at 68.

270. Epstein, *supra* note 213, at 629–32; Cole, *supra* note 42, at 1046–48, 1047 n.334.

271. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593 (1993).

272. See, e.g., *United States v. Mitchell*, 365 F.3d 215, 239–41 (3d Cir. 2004); *United States v. Sanchez-Birruetta*, 128 Fed. Appx. 571, 573 (9th Cir. 2005).

273. *Sanchez-Birruetta*, 128 Fed. Appx. at 573; Sangero & Halpert, *supra* note 59, at 70–71.

274. *Mitchell*, 365 F.3d at 225; Sangero & Halpert, *supra* note 59, at 70.

yielded 2.5 billion comparisons (50,000 × 50,000),²⁷⁵ which some courts considered as evidence that false positives cannot occur in fingerprint comparison.²⁷⁶

The main methodological flaw in the FBI “study” was the comparison of the fingerprint images against *themselves*.²⁷⁷ For the study to have been scientifically valid, the 50,000 images should have been compared with 50,000 *other* images of the same fingerprints, i.e. there should have been two different images of each fingerprint. When 50,000 images are compared against themselves, there is of course no possibility of error. The image is stored on the computer as a digital file, which is a collection of digits. When two images are identical, the digits that represent them will also be identical. Two different digital images, however, even of the same fingerprint, will be represented in the computer’s memory by different digits, and in this case, an error *is* possible. If we take what occurs in the reality of a forensic fingerprint comparison, two different images are in fact compared: the one from the crime scene and the image of the suspect’s fingerprint. Given this grave analytical mistake, this “study” has unsurprisingly drawn harsh criticism from experts in the field.²⁷⁸

Mnookin has described three approaches in American case law to fingerprint evidence.²⁷⁹ The first is to simply ignore the problem: courts hold that fingerprint evidence easily passes the *Daubert* test and disregard all the above-mentioned difficulties with this evidence—the lack of scientific testing, the lack of a meaningful error rate, and the lack of a statistical foundation or validated, objective criteria for determining a match.²⁸⁰ The second approach

275. *Mitchell*, 365 F.3d at 225. In a second experiment, they compared partial prints. *Id.* at 226.

276. *Sanchez-Birruetta*, 128 Fed. Appx. at 571, 573.

277. Sangero & Halpert, *supra* note 59, at 70.

278. Champod & Evett, *supra* note 234, at 112; David H. Kaye, *Questioning a Courtroom Proof of the Uniqueness of Fingerprints*, 71 INT’L STAT. REV. 521, 521 (2003); Pankanti et al., *supra* note 217, at 1015; Sangero & Halpert, *supra* note 59, at 70.

279. Mnookin, *supra* note 30, at 1241–65.

280. *Id.* at 1243–47. The example is the verdict in *United States v. Havvard*, 117 F. Supp. 2d 848, 855 (S.D. Ind. 2000).

seemingly applies the *Daubert* test to the fingerprint evidence but too easily concludes its admissibility: “these courts, though squirming a bit and acknowledging some of the legitimate concerns regarding the research basis for this evidence, find that, on balance, the evidence still warrants admission in its traditional form, though without fully explaining what justifies this conclusion.”²⁸¹ The third approach courts take is to allow the fingerprints expert to testify on the similarities and differences in the patterns at issue, while prohibiting or limiting her from presenting conclusions regarding the meaning of the similarities.²⁸²

Based on the 2009 NAS Report, discussed in detail below, Mnookin offers another approach, namely “Exclusion (for Now)”:²⁸³

[A]t present, pattern identification evidence does not have the empirical data to back up the claims made in court. Moreover, just as with DNA evidence—which, after an initial honeymoon period, was excluded by a number of jurisdictions for a short period of time because of concerns about the subjectivity of standards for determining a match; insufficient research into the underlying questions of population genetics; and general technical sloppiness—exclusion would be a great motivator for pursuing the research necessary to justify admissibility . . .

. . . Good proficiency tests, which show the extent to which examiners make errors in a variety of different levels of difficulty, should suffice to support a finding of adequate validity, presuming that the error rates discovered through this testing process are tolerably low, and the match between what was tested and the “task at hand” in the

281. Mnookin, *supra* note 30, at 1248–52. The example is the verdict in *United States v. Sullivan*, 246 F. Supp. 2d 700, 702–04 (E.D. Ky. 2003).

282. Mnookin, *supra* note 30, at 1252–64. An example is the verdict in *United States v. Llera Plaza*, 179 F. Supp. 2d 492, 516 (E.D. Pa. 2002).

283. Mnookin, *supra* note 30, at 1265.

particular case is sufficiently close . . .

. . . [T]he courts should care less about the details of the method . . . and more about what evidence there is to support the conclusion that the methods actually work . . .

. . .

. . . [H]ow accurate are examiners when matching latent prints to a particular source; latent prints which are often partial, frequently smudged, and perhaps even distorted?²⁸⁴

Mnookin estimates that if judges require that experts provide the error rate of their work as a prerequisite for the admissibility of fingerprint evidence, research will be conducted and very important knowledge produced.²⁸⁵ Experts should not be able to simply claim an error rate of zero and they must give up the claim that they are able to individualize.²⁸⁶ I find this view convincing and an important step on the way to safety in fingerprint evidence. In fact, in a promising decision from 2007, a court excluded fingerprint evidence, describing it as “a subjective, untested, unverifiable identification procedure that purports to be infallible.”²⁸⁷

4. *The 2009 NAS Report*

As opposed to the “50K study,” the most informative document on the accuracy of fingerprint comparison is the 2009 NAS Report on forensic science in the United States.²⁸⁸ The report deals with fingerprints under the category of “friction ridge analysis”: palm

284. *Id.* at 1265–67.

285. *Id.* at 1243.

286. *Id.* at 1275.

287. *Maryland v. Rose*, No. K06-0545, slip op. at 31 (Balt. Cty. Cir. Ct. Oct. 19, 2007). The decision was overturned by a federal judge in *United States v. Rose*, 672 F. Supp. 2d 723, 726 (D. Md. 2009).

288. NAS-2009 Report, *supra* note 3, at 2–3.

prints and sole prints.²⁸⁹ The report found that the “training of personnel to perform latent print identifications varies from agency to agency” and can amount to only a one-week-long course.²⁹⁰ “Not all agencies require [that their staff] acquire and maintain certification.”²⁹¹ As mentioned above, the technique used to examine prints is Analysis-Comparison-Evaluation-Verification (ACE-V).²⁹² In the analysis stage, the examiner considers the following features: condition of the skin, type of residue, mechanics of touch, nature of the surface touched, development technique, capture technique, and size of the latent print.²⁹³ In the next stage, a visual comparison is made between the latent print and the known print derived from the suspect.²⁹⁴ The examiner then performs source determination by evaluating whether there is identification.²⁹⁵ Last, there is verification of the first examiner’s findings by another qualified examiner, who repeats the observations and comes to a conclusion, although he or she may be aware of the conclusion arrived at by the first examiner.²⁹⁶ The NAS Report describes the process and its problematic subjective aspects as follows:

Note that the ACE-V method does not specify particular measurements or a standard test protocol, and examiners must make subjective assessments throughout. In the United States, the threshold for making a source identification is deliberately kept subjective, so that the examiner can take into account both the quantity and quality of comparable details. As a result, the outcome of a friction ridge analysis is not necessarily repeatable from examiner to examiner. In fact, recent research by Dror has

289. *Id.* at 136–55.

290. *Id.* at 136.

291. *Id.* at 137.

292. *Id.*

293. *Id.* at 137–38.

294. NAS-2009 Report, *supra* note 3, at 138.

295. *Id.*

296. *Id.*

shown that experienced examiners do not necessarily agree with even their own past conclusions when the examination is presented in a different context some time later.²⁹⁷

It is important to stress again in this context that the experts usually work with the police and have knowledge of details of the investigation, such as the fact that the suspect was identified by the victim. This extraneous knowledge is likely to influence the expert's subjective evaluations as to a match between the prints. The report notes further on this issue:

This subjectivity is intrinsic to friction ridge analysis, as can be seen when comparing it with DNA analysis . . .

. . . By contrast, before examining two fingerprints, one cannot say a priori which features should be compared . . . For these reasons, population statistics for fingerprints have not been developed, and friction ridge analysis relies on subjective judgments by the examiner. Little research has been directed toward developing population statistics, although more would be feasible.²⁹⁸

A safety approach, however, would lead to this much-needed research, which should lead in turn to the design and implementation of objective standards.

On the matter of “methods of interpretation,” the report found that:

The clarity of the prints being compared is a major underlying factor . . . Clearly, the reliability of the ACE-V process could be improved if specific measurement criteria were defined. Those criteria become increasingly important when working with latent prints that are smudged and

297. *Id.* at 139. The report refers to Dror & Charlton, *supra* note 262, at 600–16.

298. NAS-2009 Report, *supra* note 3, at 139–40.

incomplete, or when comparing impressions from two individuals whose prints are unusually similar.²⁹⁹

The report also referred to the reporting of results: “the friction ridge community actively discourages its members from testifying in terms of the probability of a match.”³⁰⁰

The report concurred³⁰¹ with Mnookin’s observations, which she stated as follows:

Experts therefore make only what they term “positive” or “absolute” identifications—essentially making the claim that they have matched the latent print to the one and only person in the entire world whose fingerprint could have produced it . . . [S]uch claims . . . are unjustified . . . Therefore, in order to pass scrutiny under *Daubert*, fingerprint identification experts should exhibit a greater degree of epistemological humility. Claims of “absolute” and “positive” identification should be replaced by more modest claims about the meaning and significance of a “match.”³⁰²

In its “summary assessment,” the report referred to Lyn Haber and Ralph Norman Haber’s work,³⁰³ where they showed that there is no scientific evidence of the validity of the ACE-V method.³⁰⁴ Examiners differ at each stage of the method in their conclusions, and no single protocol has been officially accepted by the profession; therefore, the validity of the ACE-V method cannot be tested. The report also noted that two legal decisions have highlighted the crucial

299. *Id.* at 140.

300. *Id.* at 141.

301. *Id.* at 142.

302. Jennifer L. Mnookin, *The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate*, 7 L. PROBABILITY & RISK 127, 139 (2008).

303. NAS-2009 Report, *supra* note 3, at 133.

304. L. Haber & R.N. Haber, *Scientific Validation of Fingerprints Evidence Under Daubert*, 7 L. PROBABILITY & RISK 87, 87 (2008).

issues of the lack of documentation and lack of data as to the error rate.³⁰⁵ Another justified “criticism of the latent print community is that examiners can too easily explain a ‘difference’ as an ‘acceptable distortion’ in order to make an identification.”³⁰⁶

Finally, the report made a very important recommendation: the establishment of an independent federal agency to regulate, supervise, and improve forensic sciences, to be known as the National Institute for Forensic Science (NIFS).³⁰⁷

5. *The 2012 Expert Working Group Report*

Perhaps the most significant development in this field of late is the 2012 National Institute of Standards and Technology report, entitled *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*.³⁰⁸ The report’s most noteworthy recommendation, consistent with the recommendations in the 2009 NAS Report, is as follows: “Because empirical evidence and statistical reasoning do not support a source attribution to the exclusion of all other individuals in the world, latent print examiners should not report or testify, directly or by implication, to a source attribution to the exclusion of all others in the world.”³⁰⁹ Another important recommendation relates to the problematic and flawed use of AFIS searches, as illustrated by the Brandon Mayfield case:

When comparing latent prints to exemplars generated through AFIS searches, examiners must recognize the

305. *New Hampshire v. Langill*, 945 A.2d 1, 11–12 (N.H. 2008); *Maryland v. Rose*, No. K06-0545, slip op. at 25–26 (Balt. Cty. Cir. Ct. Oct. 19, 2007).

306. NAS-2009 Report, *supra* note 3, at 145.

307. *Id.* at 189.

308. See generally NAT’L INST. OF STANDARDS & TECH., *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* (2012), <http://www.nist.gov/oles/upload/latent.pdf> [<http://perma.cc/G299-SJBP>].

309. *Id.* at 197.

possibility and dangers of incidental similarity. Adjustments such as a higher decision threshold, stricter tolerances for differences in appearance, and explicit feature weighting need to be considered. Modified quality assurance practices for this scenario also should be considered.³¹⁰

Other important recommendations in the report relate to reporting, documentation, and testimony, including the adoption of codes of ethics.³¹¹ There is also a set of recommendations that stem from a safety approach, as proposed in this article. As explained in the report:

Supervision of the staff members and management of the facilities are essential to risk reduction and quality assurance and control. Effective management requires good information about the incidence and sources of errors. Making the information available requires a culture in which both management and staff understand that openness about errors is not necessarily a path to punitive sanctions.³¹²

Also in line with the safety approach I advocate adopting are the report's recommendations to improve staff training and education,³¹³ to provide proper facilities and equipment,³¹⁴ and to channel federal support to research efforts.³¹⁵

6. *The 2016 Report to the President*

The report found that:

310. *Id.* at 199.

311. *Id.* at 200–01.

312. *Id.* at 201.

313. *Id.* at 202–03.

314. NAT'L INST. OF STANDARDS & TECH, *supra* note 308, at 203.

315. *Id.* at 203–06.

[L]atent fingerprint analysis is a foundationally valid subjective methodology—albeit with a false positive rate that is substantial and is likely to be higher than expected by many jurors based on longstanding claims about the infallibility of fingerprint analysis. The false-positive rate could be as high as 1 error in 306 cases based on the FBI study and 1 error in 18 cases based on a study by another crime laboratory. In reporting results of latent-fingerprint examination, it is important to state the false-positive rates based on properly designed validation studies.³¹⁶

The report refers to confirmation bias, contextual bias, and proficiency testing and recommends continuing efforts to improve the state of latent-print analysis.³¹⁷

Finally,

[a] second—and more important—direction is to convert latent-print analysis from a subjective method to an objective method. The past decade has seen extraordinary advances in automated image analysis based on machine learning and other approaches—leading to dramatic improvements in such tasks as face recognition and the interpretation of medical images. This progress holds promise of making fully automated latent fingerprint analysis possible in the near future. There have already been initial steps in this direction, both in academia and industry.³¹⁸

316. EXEC. OFFICE OF THE PRESIDENT, *supra* note 130, at 9–10.

317. *Id.* at 10.

318. *Id.* at 10–11.

7. *Summary and Recommendations*

It is important to stress that I do not advocate generally ruling out the admissibility of fingerprint evidence.³¹⁹ Such a sweeping move, in my view,³²⁰ would be decidedly misguided. This is indisputably significant and weighty evidence³²¹ that the courts should be allowed to continue to rely on. What is necessary, however, is that this type of evidence be improved and upgraded, and that it not be allowed to constitute the sole basis of a conviction.

In the current state of affairs, as described, there is essentially a lack of scientific grounding for fingerprint comparisons, because the possibility of a random match has never been refuted. More significantly, courts are not presented with testifying experts' error rates. It is therefore almost shocking that this evidence plays such a main role in criminal trials and convictions. Even if the Random Match Probability were proven to be very low (or even zero) and the error rate proven to be very low (but non-zero), a conviction beyond a reasonable doubt would still be impossible to establish based solely on this evidence.³²² Moreover, under the modern safety approach proposed in this article, it is necessary to strive to gather and present the most accurate evidence possible in a criminal investigation. Thus, there should be a rule requiring that examiners work for an independent federal agency rather than under the authority of the law enforcement agencies. As I will explain below, forensic science evidence—including fingerprint comparisons—should not be admissible in court unless the evidence has been developed as a

319. I have stressed this point elsewhere, along with Dr. Halpert, for example in Sangero & Halpert, *supra* note 59, at 71.

320. See Epstein, *supra* note 213, at 624–25, 627 (discussing claims that fingerprint evidence does not meet the criteria of the *Daubert* ruling).

321. Computerized models have been developed in the field of pattern recognition for the computerized comparison of fingerprints. See Joaquin Gonzalez-Rodriguez et al., *Bayesian Analysis of Fingerprint, Face and Signature Evidences with Automatic Biometric Systems*, 155 FORENSIC SCI. INT'L 126, 132–34 (2005). These models provide data on the error rate in various categories of test conditions for different systems. *Id.* at 133–34. Thus, computerized fingerprint evidence has an identification capability and enables the determination of an identification error rate. See Sangero & Halpert, *supra* note 59, at 71. There is still much to be done in this field, however.

322. See *supra* Part II.C (demonstrating the significance of even a low error rate).

safety-critical system. A substantial leap in this direction would be to adopt the recommendations made in the 2009 NAS Report, the 2012 Expert Working Group report, and the 2016 Report to the President, which conducted research to establish probabilities rather than absolute results, research on error rates, and development and implementation of appropriate protocols for fingerprint comparisons.³²³

Finally, we should also recall the possibility that a wrongdoer or police officer may transfer an innocent person's fingerprints to a crime scene to frame the innocent person. This possibility contributes to the lack of safety regarding fingerprint evidence.³²⁴

F. "Junk Science" As Evidence

During the early modern period of history from 1450 to 1750, the infamous "Satanic witch trials" were conducted in England, in which women were accused of witchcraft and devil worship.³²⁵ To prove the guilt of the accused, so-called experts were called upon, who searched—and often found—the "Devil's mark" on the women's bodies.³²⁶ In particular, they searched for the remains of what was known as the "witches' teat" by which the women purportedly nourished the Devil.³²⁷ These experts developed special methods for examining the women's bodies for these marks.³²⁸ The English courts admitted their testimony as evidence proving guilt.³²⁹ Moreover, even when the Devil's mark could not be found on a woman's body, this

323. NAS-2009 Report, *supra* note 3, at 31–32 ("Recommendation 12: Congress should authorize and appropriate funds for the National Institute of Forensic Science (NIFS) to launch a new broad-based effort to achieve nationwide fingerprint data interoperability. To that end, NIFS should convene a task force comprising relevant experts from the National Institute of Standards and Technology and the major law enforcement agencies . . . and industry, as appropriate.").

324. For a discussion of such a case in Israel of transferring a fingerprint of the suspect to the crime scene by a police officer, see SANGERO, *supra* note 8, at 190–91.

325. ORNA ALYAGON DARR, MARKS OF AN ABSOLUTE WITCH: EVIDENTIARY DILEMMAS IN EARLY MODERN ENGLAND 6–8 (2011).

326. *Id.* at 114.

327. *Id.* at 114–15.

328. *Id.* at 119.

329. *Id.* at 61.

was not regarded as a sign of innocence.³³⁰ The convicted women were sentenced to death.³³¹ The question that arises is: How far have we advanced since then? To this very day, courts unfortunately continue to admit certain types of evidence that are deserving of the title “junk science”³³²—such as microscopical hair comparisons, shoeprint comparisons, and voiceprint identification—as evidence.³³³

The 2009 NAS Report reviewed in detail the various areas of microscopical comparison and comparisons based on an expert’s subjective impression, and determined that, with the sole exception of DNA comparisons,³³⁴ none of these fields is currently grounded in science.³³⁵

The law is not a science—certainly not an exact one. In the law, crucial decisions are made in conditions of uncertainty. In scientific research, there is no need to reach a conclusion at a particular given moment and the research can be continued until it reaches an advanced stage in which precise conclusions can be made. In a trial, there is a need to arrive at a determination within a reasonable period of time, and it is not possible to wait interminably for more data and information. Science is considered precise, and therefore, it is no

330. *Id.* at 118.

331. DARR, *supra* note 325, at 73; SANGERO, *supra* note 8, at 125.

332. Even though the expression “junk science” was already in use in the 1980s, it only achieved wider recognition in the legal world following the release of the book PETER W. HUBER, *GALILEO’S REVENGE: JUNK SCIENCE IN THE COURTROOM 2* (1991), and the similarly entitled article, Peter W. Huber, *Junk Science in the Courtroom*, 26 VAL. U. L. REV. 723, 723 (1992). At a certain point, the term came to be used also to describe forensic science that leads to many false convictions. *See, e.g.*, David Bernstein, *Junk Science in the United States and the Commonwealth*, 21 YALE J. INT’L L. 123, 124 (1996); Paul C. Giannelli, *Junk Science: The Criminal Cases*, 84 J. CRIM. L. & CRIMINOLOGY 105, 105 (1993); Thomas R. May, *Fire Pattern Analysis, Junk Science, Old Wives Tales, and Ipse Dixit: Emerging Forensic 3D Imaging Technologies to the Rescue?*, 16 RICH. J.L. & TECH. 13, 13 (2010); Sangero & Halpert, *supra* note 11, at 427.

333. BARRY SCHECK ET AL., *ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED* 158–71 (2000); Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 56 (1998); Garrett & Neufeld, *supra* note 38, at 71–75; Saks & Koehler, *Forensic Identification Science*, *supra* note 35, at 892–93; GARRETT, *supra* note 7, at 95–100, 105–06.

334. NAS-2009 Report, *supra* note 3, at 7 (“With the exception of nuclear DNA analysis, however, 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.”).

335. NAS-2009 Report, *supra* note 3, at 161; SANGERO, *supra* note 8, at 127.

wonder that legal practitioners tend to pursue it. When police investigators, prosecutors, and judges base a defendant's guilt on scientific evidence they feel more secure and convinced. When a piece of evidence is truly scientific—that is, well grounded in valid and reliable scientific research—their reliance on this evidence should be commended, so long as they are not blinded into putting everything else aside and according this evidence more weight than it warrants. Even the strongest scientific evidence today—DNA genetic comparisons and fingerprint comparisons—are not sufficiently strong to serve as the sole basis for proving guilt beyond a reasonable doubt. This deficiency stems somewhat from the possibility of a random match, but has more to do with the far-more-probable likelihood of a lab error or error in an expert's analysis.³³⁶

This section is devoted to those types of evidence that are deceptively presented as “scientific” when they in fact lack a sufficient scientific basis, and moreover, when courts sometimes refer to them in their decisions as “scientific.” Junk science disguised as true science is likely to mislead judges and jurors into thinking it is actually scientific evidence and thus result in false convictions.³³⁷ In this section, I will consider additional factors that contribute to courts' misguided admission of junk science as evidence: namely, the “Sherlock Holmes myth” and “CSI effect,” and the misleading or erroneous presentation of evidence by police and prosecution experts.

Mention of an additional issue—beyond the existence of junk science—affecting the forensic sciences field, such as DNA comparisons, is necessary here. Namely, an uneven balance of power exists between the prosecution and defense such that almost all the scientific evidence submitted in court is presented by the prosecution.³³⁸ Even on the few occasions on which the defense

336. SANGERO, *supra* note 8, at 128; Sangero & Halpert, *supra* note 59, at 61.

337. SANGERO, *supra* note 8, at 129.

338. An empirical study conducted in the United States found that court-appointed defense attorneys hire experts in only 2% of their criminal cases and in only 17% of their manslaughter cases. See Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 WIS. L. REV. 291, 333 (2006) (citing Michael McConville & Chestler L. Mirsky, *Criminal Defense of the Poor in New York City*, 15 N.Y.U. REV. L. & SOC. CHANGE 581, 764 (1986–1987)); see also Sangero &

counsel submits expert testimony, judges tend to prefer the testimony of the prosecution's expert witnesses.³³⁹

Crime laboratories, which engage in what is called “forensic science” and produce scientific evidence, are usually not autonomous, but rather operate under the direct authority of the police and prosecution.³⁴⁰ The laboratory personnel work closely and routinely with the police investigators and also with prosecutors at times, making it hard to expect them to be completely objective in performing their jobs.³⁴¹ It is reasonable to assume that police investigators bring pressure to bear on lab staff to find evidence that supports the suspect's guilt. In addition, the police investigators feed the lab workers details of the investigation that are completely irrelevant to the tests they perform, but that reinforce the perception of the suspect's guilt, such as the fact that the suspect confessed or was identified by the victim.³⁴² Here too, the misconception of the suspect's guilt (“Tunnel Vision”) plays a vital role: many in the law enforcement system tend to assume that the suspect is guilty and that they need simply find evidence that proves this guilt.³⁴³ At times, the lab workers are requested, by definition, to conduct only the tests that are likely to incriminate the suspect, without performing any other tests that could prove his innocence or even incriminate someone else.³⁴⁴ At a later stage, the testimony of the expert is also likely to be misleading. Thus, for example, experts tend to testify to a significantly higher level of precision in the test that they performed than the real degree of precision according to the most up-to-date scientific research.³⁴⁵

Halpert, *supra* note 11, at 434.

339. Garrett & Neufeld, *supra* note 38, at 90–91.

340. *Id.* at 33–34; DiFonzo, *supra* note 25, at 4–5.

341. Garrett & Neufeld, *supra* note 38, at 33–34.

342. Thompson, *supra* note 30, at 1034.

343. Findley & Scott, *supra* note 338, at 292 (“[F]ocus on a suspect, select and filter the evidence that will ‘build a case’ for conviction, while ignoring or suppressing evidence that points away from guilt.”).

344. Garrett & Neufeld, *supra* note 38, at 81–83.

345. *See id.* at 84.

Many judges are unaware of the possibility—which always exists—of a lab testing error and do not question the experts on this; the experts, for their part, often do not bother to volunteer such information. Moreover, with regard to the majority of the branches of “forensic science,” even if the experts wanted to provide the court with information about the precision of the tests and their estimated error rate, they would not be able to do so due to the lack of existing research and data.³⁴⁶

In addition, the majority of the “forensic sciences” branches—particularly those in which experts testify on the comparisons they perform between marks at the crime scene and marks made by an object or limb belonging to the defendant—have been developed especially for the purpose of solving crimes; they were not developed by scientists.³⁴⁷ These branches are not, therefore, based on methodical scientific research with an adequate database, but rather on experience gathered by “experts” through their use of the very system developed to solve crimes.³⁴⁸

It has been suggested that a distinction be drawn between scientists and technicians.³⁴⁹ To guarantee objectivity, scientists use “blind tests” to determine whether a particular result is correct or the product of contamination. In contrast, technicians usually know how but not why. Many forensic fields fall into the sphere of technicians’ work and not scientific work. Technicians, even when they are doing their job properly, lack the necessary scientific training to plan experiments that will turn their work into science. In addition, they lack the necessary databases for conducting statistical calculations. In fact, many forensic areas were developed by law enforcement agents, who, in their attempts to solve crimes and, often, to find incriminating evidence against the suspect, tended to turn unfounded

346. SANGERO, *supra* note 8, at 106.

347. *Id.*

348. *Id.* at 107; *see generally* INNOCENCE PROJECT, www.innocenceproject.org [<https://perma.cc/L9W8-TVS2>] (last visited July 7, 2018).

349. Craig M. Cooley, *Forensic Science or Forgettable Science?*, 80 IND. L.J. 80, 81 (2005); SANGERO, *supra* note 8, at 107.

premises of the individualization of certain crime scene marks into so-called science. Allegedly, if a mark found at the crime scene resembles the mark left by an object or limb belonging to the defendant, this necessitates the conclusion that the defendant's object or limb left the mark found at the crime scene. This is the case with regard to microscopical comparisons of hair, fibers, tool marks, weapon marks, shoeprints, teeth prints, and even ear prints.³⁵⁰ The possibilities are limitless.

At the suggestion of Michael Risinger, we should also include the great detective Sherlock Holmes on the list of those responsible for this current state of shoddiness.³⁵¹ What do we all love about Sherlock Holmes? Among other things, he has contributed to the great faith we all have in the brain's ability to perpetually and precisely deduce the criminal's identity based on how the criminal acted from the evidence found at the crime scene. Not only the public at large but also forensic practitioners have apparently been raised on the Sherlock Holmes myth. Similarly, many wrongly think that the work of forensic labs is as perfect as it appears on the popular *CSI* television series, whose influence on jurors and other entities in the criminal justice system has been significant and harmful.³⁵²

Studies have uncovered the development of a nonscientific subculture in laboratories, developed out of an eagerness to please police investigators. This subculture seeks to "deliver the goods" to convince the court rather than adhere to science, even if the results are equivocal and are likely to disappoint those who believe that the suspect committed the crime.³⁵³ This was the case in the infamous Houston crime lab scandal: the lab was closed down after it emerged

350. BEATRICE SCHIFFER & CHRISTOPHE CHAMPOD, *Judicial Error and Forensic Science: Pondering the Contribution of DNA Evidence*, in *WRONGFUL CONVICTION: INTERNATIONAL PERSPECTIVES ON MISCARRIAGES OF JUSTICE* 33, 38 (C. Ronald Huff & Martin Killias eds., 2008); SANGERO, *supra* note 8, at 107.

351. Risinger, *supra* note 27, at 527; SANGERO, *supra* note 8, at 108.

352. NAS-2009 Report, *supra* note 3, at 48; Bolden, *supra* note 11, at 425; Difonzo, *supra* note 25, at 2–3 (noting the "'CSI Effect,' popularly defined as 'the perception of the near-infallibility of forensic science in response to the TV show'"); Mnookin, *supra* note 30, at 1209.

353. Saks & Koehler, *Forensic Identification Science*, *supra* note 35, at 893; SANGERO, *supra* note 8, at 108.

that for many years, it had been systematically providing erroneous lab results, which had served as the basis for many convictions.³⁵⁴ The exposure of the scandal in 2002 on a television program left authorities no choice but to make a sweeping investigation of the matter.³⁵⁵ After some partial reports, a comprehensive report was finally released in 2007 regarding the laboratory's operation.³⁵⁶ The investigation, which reviewed over a thousand cases, uncovered wrongdoing by nine different crime lab workers.³⁵⁷ The investigation report pointed to a long list of malfunctions in the police lab's work, including: failing to perform appropriate control experiments in DNA testing; systematically misleading reporting regarding the statistical significance of DNA matches that were found; failing to report possibly exculpatory findings in suspects' favor; experts' misrepresenting and exaggerating their credentials and training in court; misleading representation of findings; and even fabricating findings.³⁵⁸ It was found that these practices had continued for an entire decade, until exposed on the television program.³⁵⁹ Similar malfunctioning was exposed at the Illinois state police crime lab and in the work of Texas forensic pathologist Ralph Erdmann, whose testimony contributed to at least twenty death penalty convictions and who was convicted of falsifying autopsies.³⁶⁰

One central problem is that some of the types of evidence submitted in the past to the court—and, in some cases, that continue to be presented—are based on forensic methods and techniques that are entirely unreliable, such as microscopical comparisons of hair, bite marks on the skin, shoeprints, and voiceprints.³⁶¹ As clarified in the 2009 NAS Report, and as many scientists have cautioned in their

354. MICHAEL BROMWICH ET AL., FINAL REPORT OF THE INDEPENDENT INVESTIGATOR FOR THE HOUSTON POLICE DEPARTMENT CRIME LABORATORY AND PROPERTY ROOM 5 (June 13, 2007), <http://www.hpdlabinvestigation.org/reports/070613report.pdf> [<https://perma.cc/2SN4-BKYR>].

355. *Id.* at 4.

356. *Id.* at 1; *see also* SANGERO, *supra* note 8, at 108.

357. Bolden, *supra* note 11, at 418–19.

358. *Id.* at 418.

359. Thompson, *supra* note 30, at 1037.

360. Bolden, *supra* note 11, at 418.

361. NAS-2009 Report, *supra* note 3, at 3.

work, these identification methods have no scientific grounding, are not based on data, and are unreliable.³⁶²

A second central problem is that the data collected in the framework of the Innocence Project revealed that the majority (61%) of expert testimony for the prosecution in cases with false convictions was invalid and faulty; the experts had presented the findings of the tests that they had performed misleadingly and erroneously and had arrived at conclusions in an unscientific manner.³⁶³ This was the case with regard to both the unreliable types of evidence discussed above as well as reliable scientific evidence, such as DNA comparisons and serological blood type tests.³⁶⁴ Common to the experts' errors in their testimonies was their biased presentation of the forensic evidence as stronger than it truly was in reality.³⁶⁵

According to *Daubert*, judges are supposed to be “gatekeepers,” preventing entry into the courtroom of evidence that is not scientifically reliable or valid. However, in reality, usually only the prosecution succeeds in submitting an expert opinion—which is most often incriminating—because defendants lack the necessary resources to submit expert opinions in their favor. Because judges do not have expertise in all of the scientific or pseudoscientific fields presented to them, they are often misled by the prosecution's biased expert opinion, whereas the defense attorney lacks the tools for refuting it. It is interesting to note that in civil trials, especially tort lawsuits—which never involve capital cases or human liberty and tend to revolve solely around matters of money—judges delve far more deeply into the evidence presented as scientific, and in practice, actually function as gatekeepers and prevent junk science from being admitted at trial. How does this happen? In civil law, expert opinions are often submitted by both parties to the litigation—the plaintiff and the defendant—which allows the judge to compare and confront the

362. *Id.* at 161; SANGERO, *supra* note 8, at 135.

363. GARRETT, *supra* note 7, at 90.

364. *Id.*

365. *See id.*; *see also* Garrett & Neufeld, *supra* note 38, at 20; SANGERO, *supra* note 8, at 135–36.

evidence and make a far deeper investigation into its reliability and validity.³⁶⁶ In criminal law, judges tend to rely on the prosecution's experts' guarantees that they perform their work in line with precise scientific standards, and tend to rely on the defense attorneys to expose any imprecision in that work. Both these assumptions are misplaced.³⁶⁷

An additional problem is that the professional expertise in the fields examined in crime laboratories is acquired in the framework of entities associated with the police and prosecution. The experts who testify at criminal trials are almost always prosecution witnesses and almost always work for the police. In effect, they see themselves as part of the law enforcement system and consider their job to be assisting law enforcement in finding evidence pointing to the suspect's guilt. Here, again, we witness the destructive effect of preconceptions of the suspect's guilt.³⁶⁸

An extreme example of such an “accomplished” expert was Fred Zain, the head of the West Virginia state crime lab. The Innocence Project's work revealed that Zain would often falsify the results of tests to fabricate supposedly scientific evidence that would incriminate suspects.³⁶⁹ An investigation revealed that he had shamelessly lied in his testimonies as an expert before courts in two different states in the United States and had manipulated test results in 134 different cases over thirteen years.³⁷⁰ Moreover, Zain would often testify in court that he himself had performed the tests when in fact others had conducted them.³⁷¹

Forensic experts do not perform their examinations as “blind tests”—that is, without knowledge of additional information about

366. See Beecher-Monas, *supra* note 333, at 56; GARRETT, *supra* note 7, at 91; Kesan, *supra* note 20, at 2040.

367. Garrett & Neufeld, *supra* note 38, at 97; SANGERO, *supra* note 8, at 136.

368. SANGERO, *supra* note 8, at 136.

369. Bolden, *supra* note 11, at 317–18.

370. *Id.* at 418.

371. See SCHECK ET AL., *supra* note 333, at 107–25 (providing a more detailed description of the Fred Zain case); see also Paul C. Giannelli & Kevin C. McMunigal, *Prosecutors, Ethics, and Expert Witnesses*, 76 *FORDHAM L. REV.* 1494, 1497–98 (2008); SANGERO, *supra* note 8, at 136–37.

the investigation—and in fact, police investigators often inform them on the details of the investigation. It is only natural as human beings that they are influenced by the knowledge that additional evidence exists supporting the suspect's guilt, such as his identification by the victim, confession to the police, or even an additional piece of scientific evidence. All tests have the subjective component of the expert's assessment and interpretation of the findings. The knowledge that there are other pieces of evidence against the suspect is likely to bias the expert toward an incriminating interpretation of findings that are not unequivocal.³⁷² Moreover, in many cases, experts have been prevented from performing tests that are likely to rule out the suspect's involvement in the crime, and in other cases, the prosecution withholds from the defense findings pointing to the defendant's innocence. In their summations at trial, prosecutors often present the lab findings in a misleading and erroneous manner, so that they will be perceived as incriminating.³⁷³

Another phenomenon that has emerged is that those experts who give misleading and erroneous testimony are not merely a few rotten apples, but are rather quite numerous. According to the findings of the Innocence Project, from among the 250 first exonerations obtained in the framework of the Project, 81 different experts working in 54 different laboratories in 28 different states across the United States were found to have given faulty expert testimony.³⁷⁴ This should give one pause; consider the thousands of other trials at which these same experts testified but the convicted defendants have not been retried and, in all likelihood, will never be retried because, among other reasons, there are no genetic samples that can be tested

372. GARRETT, *supra* note 7, at 92; *see also* Dror et al., *supra* note 260, at 74 (discussing experts' biases); Itiel E. Dror et al., *When Emotions Get the Better of Us: The Effect of Contextual Top-Down Processing on Matching Fingerprints*, 19 APPLIED COGNITIVE PSYCHOL. 799, 799 (2005); Itiel E. Dror, *Cognitive Neuroscience in Forensic Science: Understanding and Utilizing the Human Element*, PHIL. TRANSACTIONS ROYAL SOC'Y B: BIOLOGICAL SCI. 370, 370 (2015); Itiel E. Dror, *A Hierarchy of Expert Performance*, 5 J. APPLIED RESEARCH IN MEMORY & COGNITION 121, 121 (2016).

373. GARRETT, *supra* note 7, at 111–13; *see also* Garrett & Neufeld, *supra* note 38, at 34; SANGERO, *supra* note 8, at 138.

374. GARRETT, *supra* note 7, at 93; *see* Garrett & Neufeld, *supra* note 38, at 23–24 (providing more detail).

in these cases.³⁷⁵ Here, again, we feel the effect of the Hidden Accidents Principle in criminal law.

Even when (genuine) scientists indicate in their research that certain methods are in no way scientific and call for an improvement of laboratory work practices, and even with methods that are grounded in science (particularly genetic comparisons), forensic scientists tend to resist the recommendations for improvement. They are used to the practices they learned in their training, have accepted them as correct, and followed them for many years; therefore, they see these suggestions for change as a personal attack that they almost instinctively attempt to fight.³⁷⁶

Yet, to reduce the number of false convictions that stem from faulty expert testimony, there is an urgent need for fundamental changes in this field. This includes separating the forensics lab work from the police work to enable the lab to conduct forensic testing in autonomous, objective labs, instituting “blind” expert checks where the experts do not know that their work is being checked, requiring that experts base their work on data and provide the courts with precise data regarding tests’ error rates and adopting additional changes recommended in the 2009 NAS Report.³⁷⁷

Finally, a technique that is particularly illustrative of junk science is microscopical hair comparison.³⁷⁸ Herman Douglas May was seventeen years old in 1988 and had been involved in a few minor offenses, such as stealing a guitar from a man who alleged owed May money but who refused to pay him.³⁷⁹ Around the time May stole the guitar, a burglary was committed during which a woman was raped.³⁸⁰ The woman identified May as the rapist in what the court

375. GARRETT, *supra* note 7, at 93; SANGERO, *supra* note 8, at 138–39.

376. Risinger, *supra* note 27, at 535; Sangero & Halpert, *supra* note 11, at 442; Mnookin et al., *supra* note 233, at 744–60.

377. NAS-2009 Report, *supra* note 3, at 82, 184–89; SANGERO, *supra* note 8, at 139.

378. Beth Albright & Debbie Davis, *Guilty Until Proven Innocent: The Case of Herman Douglas May*, 30 N. KY. L. REV. 585, 594 (2003); Sangero & Halpert, *supra* note 11, at 447.

379. Albright & Davis, *supra* note 378, at 586–87.

380. *Id.* at 587–88.

described as an unfairly suggestive identification procedure.³⁸¹ At trial, moreover, the forensic expert gave testimony against May, according to which hair found at the scene of the crime resembled May's hair, based on the expert's microscopical examination of the hair samples, and May could thus be the source of the hair at the crime scene.³⁸² May was convicted, and twelve years later was it determined that he was not the rapist, only after mitochondrial DNA testing was performed on both the sperm and hair found at the crime scene.³⁸³ A similar, infamous case is the false conviction of Gary Dotson for a rape that had never occurred, based on the microscopical comparison of pieces of hair.³⁸⁴

In twenty-one of the first seventy exonerations in which the Innocence Project was involved, experts gave erroneous testimony on the microscopical comparison of hair.³⁸⁵ In a later study, it emerged that in seventy-five of the Project's first 250 exonerations, the convictions had also been based on microscopical comparisons of hair,³⁸⁶ and in twenty-nine of these cases experts gave erroneous testimony.³⁸⁷ In eighteen of the cases, moreover, the experts had grounded their testimony on an individualization claim; that is to say, that the hair found at the scene of the crime was unique to the defendant.³⁸⁸ In six of the exoneration cases, mitochondrial DNA testing ruled out the expert's assessment of a match between the hair at the scene of the crime and the defendant's hair.³⁸⁹

381. *Id.* at 594. For more on the Supreme Court's lenient approach to suggestive identification and for criticism of this case law, see Gary L. Wells & Deah S. Quinlivan, *Suggestive Eyewitness Identification Procedures and the Supreme Court's Reliability Test in Light of Eyewitness Science: 30 Years Later*, 33 L. HUM. BEHAV. 1, 1 (2009).

382. See Albright & Davis, *supra* note 378, at 592.

383. *Id.* at 599; Sangero & Halpert, *supra* note 11, at 447.

384. See GARRETT, *supra* note 7, at 84–89. At a later stage, the young woman confessed to having fabricated the rape and falsely accused Dotson in an attempt to hide from her parents that she had had consensual sexual relations with her boyfriend (who was not Dotson). She also described the events in a book she wrote, CATHLEEN C. WEBB & MARIE CHAPIAN, *FORGIVE ME* (1985).

385. Albright & Davis, *supra* note 378, at 592; Sangero & Halpert, *supra* note 11, at 447–48.

386. GARRETT, *supra* note 7, at 90.

387. *Id.*

388. Garrett & Neufeld, *supra* note 38, at 47.

389. *Id.* at 51.

The method by which hair is microscopically compared has not changed much over the last century.³⁹⁰ Hair (at times only a sole strand) found at the crime scene is compared to hair taken from the suspect.³⁹¹ Routine practice is to take fifty strands of hair from the suspect when comparing head hair, and twenty-five pieces of hair when comparing the hair from another part of the body.³⁹² In the first stage of the process, the expert makes a number of determinations with regard to the hair found at the crime scene: whether it is indeed hair and not some other fiber; whether it is human hair or animal hair; the part of the body from which the hair comes; the race of the person to whom the hair belongs; whether the hair has been dyed; whether the hair fell out naturally at the crime scene or was forcibly torn out; and whether the hair was cut.³⁹³ In the second stage, the expert examines the hair without using any instrument to determine its color and structure (straight, wavy, or curly).³⁹⁴ In the third stage, the expert examines the hair with a microscope and determines a set of characteristics that are then compared to the same characteristics of the suspect's hair.³⁹⁵ These characteristics relate to hair color, structure, the structure of the hair follicle, and acquired features, such as cosmetic treatments or flaws in the hair.³⁹⁶

However, these characteristics are not consistent for even one individual's hair. And as they can vary on one person, there is a broad overlap in the characteristics of the hair of different people. Therefore, this type of comparison between pieces of hair from two different people often points to an alleged match. In addition, as this

390. Walter F. Rowe, *The Current Status of Microscopical Hair Comparisons*, 1 SCI. WORLD 868, 869 (2001); Sangero & Halpert, *supra* note 11, at 448.

391. NAS-2009 Report, *supra* note 3, at 157.

392. *Id.*

393. Paul C. Giannelli, *Microscopic Hair Comparisons: A Cautionary Tale*, 46 CRIM. L. BULL., no. 3, 2010, art. 7, at 1–2.

394. *Id.*

395. *Id.*

396. Rowe, *supra* note 390. For more details, see SCI. WORKING GRP. ON MATERIALS ANALYSIS (SWGMAT), FORENSIC HUMAN HAIR EXAMINATION GUIDELINES 10.1–10.5 (2005) https://www.nist.gov/sites/default/files/documents/2016/09/22/forensic_human_hair_examination_guidelines.pdf [<https://perma.cc/GVY8-9TXN>] [hereinafter SWGMAT]; Sangero & Halpert, *supra* note 11, at 448.

comparison entails subjective determinations, different experts are likely to arrive, and do arrive, at conflicting conclusions.³⁹⁷ There is no possibility of determining individuality based on a microscopical comparison of pieces of hair, and the probabilistic strength of such evidence is unknown. On this subject, the National Academy of Sciences stated: “No scientifically accepted statistics exist about the frequency with which particular characteristics of hair are distributed in the population. There appear to be no uniform standards on the number of features on which hairs must agree before an examiner may declare a “match.””³⁹⁸ The same report further stated that “in cases where there seems to be a morphological match (based on microscopic examination), it must be confirmed using mtDNA analysis, microscopic studies alone are of limited probative value. The committee found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA.”³⁹⁹

Today, a genetic comparison of hair can be performed, which is very precise and can be used to test the (weak) strength of the microscopic comparisons conducted in the past. There are two types of genetic tests.⁴⁰⁰ The first tests the DNA found in the cell in the root of the hair.⁴⁰¹ This test is the preferred one as, aside from identical twins, no two people share the same DNA.⁴⁰² However, the hair root is often not available for testing.⁴⁰³ In such circumstances, the second type of test can be performed: mitochondrial DNA testing.⁴⁰⁴ The working hypothesis is that mitochondrial DNA is maternally inherited.⁴⁰⁵ Were it not for the occurrence of mutations,

397. Garrett & Neufeld, *supra* note 38, at 49; Sanger & Halpert, *supra* note 11, at 448.

398. NAS-2009 Report, *supra* note 3, at 160.

399. *Id.* at 161.

400. SWGMAT, *supra* note 396, at 10.1–10.4; Sanger & Halpert, *supra* note 11, at 449.

401. SWGMAT, *supra* note 396, at 10.1–10.4; Sanger & Halpert, *supra* note 11, at 449.

402. SWGMAT, *supra* note 396, at 10.1–10.4.

403. *Id.*

404. *Id.* at 16.

405. Alice R. Isenberg & Jodi M. Moore, *Mitochondrial DNA Analysis at the FBI Laboratory*, FORENSIC SCI. COMM. 1 (1999), <http://www.fbi.gov/hq/lab/fsc/backissu/july1999/dnatext.htm#> [<http://perma.cc/B6NA-MB99>]; Richard E. Giles et al., *Maternal Inheritance of Human Mitochondrial DNA*, 77 PROC. NAT'L ACAD. SCI. 6715, 6715 (1980).

everyone would have identical mitochondrial DNA; but mutations have led to differentiations across population groups.⁴⁰⁶ Thus, mitochondrial testing can rule out the possibility that the suspect committed the crime, but cannot determine the identity of the actual perpetrator because many people share the same mitochondrial DNA, passed on to them through a shared matrilineal line.

The only circumstances in which some probative strength can be accorded to microscopical hair comparisons is when a suspect can be ruled out as the perpetrator of the crime.⁴⁰⁷ The National Academy of Sciences 2009 report stated the following regarding this possibility:

The results of analyses from hair comparisons typically are accepted as class associations; that is, a conclusion of a “match” means only that the hair could have come from any person whose hair exhibited—within some levels of measurement uncertainties—the same microscopic characteristics, but it cannot uniquely identify one person. However, this information might be sufficiently useful to “narrow the pool” by excluding certain persons as sources of the hair.⁴⁰⁸

The general consensus is that microscopical hair comparisons are junk science.⁴⁰⁹ This realization was possible due to the development of genetic comparisons, which proved conclusively that microscopical comparisons are far from being scientific. One of the lessons that should be learned from this development is the definite possibility that courts will admit nonscientific evidence as scientific evidence and even convict based on that evidence. The problem is that there are still areas in which the courts treat junk science like

406. Isenberg & Moore, *supra* note 405, at 2.

407. Giannelli, *supra* note 393, at 2 (“There is also agreement that, with sufficient exemplars, a person may be excluded as a suspect.”); Sangero & Halpert, *supra* note 11, at 451.

408. NAS-2009 Report, *supra* note 3, at 156.

409. See, e.g., SCHECK ET AL., *supra* note 333, at 161–63; Sangero & Halpert, *supra* note 11, at 451–52.

true science, and there is no possibility of providing compelling proof (through DNA testing or any other strong, concurred-upon technique) that the method is not scientific and must not be relied upon in a criminal trial. This is also the case with voice comparisons and shoeprint comparisons.⁴¹⁰

I have discussed the critical problems with junk science and elaborated on some of the changes necessary to make forensic evidence a more precise and scientific field, so as to improve the factual determinations in criminal trials and prevent judges from being misled. As an intermediate remedy—until the necessary fundamental changes are implemented—the courts must be more rigorous in examining “scientific evidence” brought before them and not admit dubious evidence warranting the label “junk science.” In line with the *Daubert* rule, for a given type of evidence to be admissible it must meet accepted scientific standards and be reliable, valid, and testable but, of course, unrefuted. In addition, in light of past experience, the courts must regard experts testifying before them with measured suspicion and not put blind faith in their testimony.

To conclude this discussion of the phenomenon of junk science and its hazards, I would like to propose a simple test that will assist judges in distinguishing between an area that is most certainly not scientific and one that could be scientific but must be more deeply examined in line with the *Daubert* rule. I call this test the MIT test (Massachusetts Institute of Technology test). I came up with this test recently while giving a lecture on “Scientific Evidence versus ‘Junk Science’” to Israeli judges at a workshop. At the end of the lecture, the judge who had organized the workshop thanked me and recounted that because I “go against the flow,” he had deliberated whether to schedule my lecture at the beginning of the week-long workshop or at the end. I responded by telling the judges that this was not the important issue, and what is actually far more relevant is that almost all of the workshop lecturers were police forensic lab

410. I elaborated on this at length in a coauthored article with Dr. Mordechai Halpert, Sangero & Halpert, *supra* note 11.

practitioners. I suggested to the judges that they learn about scientific evidence not from police lab technicians but from members of the academia: professors from the Weizmann Institute of Science or from the Technion (Israeli Institute of Technology), the Israeli counterparts of MIT.

The judge who had organized the workshop regretfully noted that research work is not being conducted at universities in each forensic field. I immediately proposed the following to the audience of judges: “This, then, is an easily applied selection test, which can assist you judges in preventing junk science from entering your courts: when a piece of so-called ‘scientific evidence’ is submitted to you but is not researched at the Weizmann Institute or Technion, this is proof that it is not science!”

I propose that American judges can apply the same test, simply substituting in MIT for the Weizmann Institute and Technion. Of course, this does not mean that any alleged expertise based on a field in which there is meaningful academic work should be automatically admitted as scientific evidence. But in the absence of academic interest and work, we can be certain that this is not science. This MIT test is an easy selection test relative to the *Daubert* rule, for it avoids both the complicated application of the *Daubert* standard and the embarrassment of the incorrect application of the tests by judges.⁴¹¹ If the courts were to apply such a test, they would never admit as scientific evidence opinions submitted by charlatan experts with regard to microscopical hair comparisons and shoeprint comparisons, nor would they convict and send to prison defendants based on other junk science evidence.

411. On the incorrect application of the *Daubert* rule by the courts, see Mnookin et al., *supra* note 233, at 758 (“[E]ven after *Daubert* . . . 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. If courts are not going to insist 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?”).

III. Safety Measures

A. Developing Forensic Science Evidence As a Safety-Critical System

Based on a single piece of DNA evidence that was obtained when a sample was run through a database, Daryl Mack was convicted of murder and sentenced to death;⁴¹² he was executed on April 26, 2006.⁴¹³ Yet, as we have seen, there is a possibility, even if low, of computer software used in the DNA comparison producing erroneous findings, just as a lab testing error is possible. This is only one context in which forensic evidence can put the life of an innocent person at risk. As Dr. Halpert and I have shown elsewhere, any device used to produce forensic evidence is fundamentally a “safety-critical system” in that it endangers human life.⁴¹⁴ Studies have demonstrated how erroneous scientific evidence occasionally leads to false convictions, and this is no less catastrophic than a car’s brake failure. Yet no mandatory regulation is in place—regarding any forensic evidence whatsoever—to supervise forensic software or device development in accordance with safety-critical standards, despite the broad implementation of such regulations in other fields involving life and death.⁴¹⁵

Elsewhere we have shown that the outdated Fly-Fix-Fly method⁴¹⁶ is not a sufficient safety system, and that “black box” testing⁴¹⁷ does not suffice either. Thus, it should be mandatory that devices producing forensic science evidence be developed using safety methods that are suited to their nature as safety-critical systems.⁴¹⁸ However, the greater hazard is that not even the Fly-Fix-Fly method has been implemented in the criminal justice system.⁴¹⁹ When the

412. Mack v. State, 75 P.3d 803, 803–04 (Nev. 2003); Halpert & Sangero, *supra* note 2, at 83.

413. Halpert & Sangero, *supra* note 2, at 83.

414. *Id.*

415. *Id.*

416. Sangero & Halpert, *supra* note 4, at 1297.

417. Halpert & Sangero, *supra* note 2, at 83–88.

418. *Id.* at 83.

419. *Id.* at 85.

manufacturer of forensic evidence equipment (such as a breathalyzer) markets a device that occasionally produces erroneous results due to a design or software defect, the chances that a court will detect the error are slim. In contrast, an airplane with a safety defect will necessarily be involved in an accident or incident (a “near miss”) at some point in time, and the defect will be exposed. This is not the case with flaws in forensic-evidence-producing devices. If an innocent defendant is convicted based on erroneous scientific evidence (assuming that the law allows a conviction to be based on a single piece of evidence), his claim of innocence will not be considered a refutation of the reliability of the forensic device, and the chances of proving a testing error are low. In addition, many view a conviction to be in and of itself confirmation of the device’s reliability. Therefore, to prevent or minimize the possibility of errors in forensic equipment, great precautions must be taken in the design and manufacturing processes. Due to the Hidden Accidents Principle, these are the only stages at which there is a reasonable possibility of discovering and avoiding defects.⁴²⁰

A forensic device developed in accordance with safety-critical standards can be expected to produce more precise and reliable scientific evidence.⁴²¹ To begin with, there will be fewer false positives; that is, fewer cases in which an innocent suspect or defendant is implicated by erroneous test results in a crime she did not commit. In addition, there will be fewer false negatives as well, in which a test or device erroneously rules out the actual perpetrator of the crime being investigated. This, of course, would lead to more efficient criminal law enforcement, making the safety improvement of forensic devices a win-win situation.

Finally, legislators should enact regulation regarding the development process of forensic equipment designed to be used by the criminal justice system.⁴²² An approval requirement should be set

420. As Dr. Halpert and I demonstrated in Halpert & Sangero, *supra* note 2, at 77–82, the special legal proceedings in *Chun* concerning the breathalyzer exemplify this well.

421. *Id.* at 89.

422. *Id.* at 93.

for manufacturers of forensic devices similar to what is required of manufacturers of medical diagnostic devices.⁴²³ Legislators should also set a rule for the admissibility of evidence produced by a forensic device in criminal proceedings requiring that the device be developed and supervised as a safety-critical system.⁴²⁴

B. Other Safety Changes in Scientific Evidence

What changes and reforms must the forensic sciences undergo to contribute to the legal field without misleading it? Important recommendations to this end were made in the 2009 NAS Report,⁴²⁵ the National Association of Criminal Defense Lawyers (NACDL) report from 2010,⁴²⁶ the Report to the President on Forensic Science in Criminal Courts from 2016,⁴²⁷ and various other research studies.⁴²⁸

- (1) The 2009 NAS Report recommended establishing a National Institute of Forensic Science (NIFS),⁴²⁹ with a similar recommendation made also in the 2010 NACDL report.⁴³⁰ This Institute could be easily integrated into the Safety in the Criminal Justice System Institute (SCJSI) when established.⁴³¹ Whether in the guise of the NIFS or the general SCJSI, such a federal agency should be tasked with the following: to set and enforce best practices for forensic science; to set standards for mandatory accreditation of forensic laboratories and mandatory certification for forensic

423. *Id.*

424. *Id.* at 93–94.

425. NAS-2009 Report, *supra* note 3, at xix.

426. NAT'L ASS'N OF CRIMINAL DEF. LAW. (NACDL), PRINCIPLES AND RECOMMENDATIONS TO STRENGTHEN FORENSIC EVIDENCE AND ITS PRESENTATION IN THE COURTROOM 1 (2010), www.nacdl.org/WorkArea/DownloadAsset.aspx?id=17775 [<https://perma.cc/555G-TDVM>] [hereinafter NACDL-2010 Report].

427. EXEC. OFFICE OF THE PRESIDENT, *supra* note 130, at 2.

428. Halpert & Sangero, *supra* note 2, at 93–94; Sangero & Halpert, *supra* note 4, at 1322.

429. NAS-2009 Report, *supra* note 3, at 19.

430. NACDL-2010 Report, *supra* note 426, at 3–4.

431. *See supra* note 1; Sangero & Halpert, *supra* note 4, at 1324.

examiners; to improve research and educational programs; to establish a standard terminology to be used in reporting and testing and model laboratories reports; and so on.⁴³² According to the 2016 Report to the President there is progress in this direction.⁴³³

- (2) A very important recommendation made in the 2009 NAS Report was achieving autonomy and objectivity by removing all public forensic laboratories from the administrative control of law enforcement agencies.⁴³⁴ In this context, it is crucial to ensure that all laboratory personnel receive only the minimal information required for performing the testing and that they be given no additional details about the suspect (optimally, they should not even know that the sample has been taken from a suspect) or about the case that are likely to bias them in performing the test.⁴³⁵
- (3) Some of the more important recommendations in the 2009 NAS Report related to research development: the promotion of scholarly, competitive peer-reviewed research regarding the validity of forensic methods and accuracy of forensic analyses and data collection, as well as research on human observer bias and sources of human error in forensic testing.⁴³⁶ In the wake of this report, thirteen different experts, from both sides of the fence—forensic science professionals and academic scholars—joined together to write an article aimed at setting a framework for research culture in the forensic sciences.⁴³⁷ Their conclusion was as follows:

We all believe that many forms of forensic science today stand on an insufficiently developed empirical

432. NAS-2009 Report, *supra* note 3, at 19–22.

433. EXEC. OFFICE OF THE PRESIDENT, *supra* note 130, at 125, 131.

434. NAS-2009 Report, *supra* note 3, at 24.

435. *See supra* notes 219–225 and accompanying text.

436. NAS-2009 Report, *supra* note 3, at 19, 22, 24.

437. Mnookin et al., *supra* note 233, at 725.

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.⁴³⁸

In response to the question, “What is a research culture?” the experts explained that it is:

[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.⁴³⁹

To this important explanation, I would add that the need for a culture of research in forensic science would be clear to all if there were a culture of safety in the criminal justice system.

- (4) Accreditation and certification: There should be mandatory laboratory accreditation and individual certification of forensic science professionals.⁴⁴⁰

438. *Id.* at 778.

439. *Id.* at 740.

440. NAS-2009 Report, *supra* note 3, at 25 (Recommendation 7).

- (5) A code of ethics: The 2009 NAS Report recommended that the NIFS establish a national code of ethics for forensic science and mechanisms for its enforcement.⁴⁴¹
- (6) Education and training: The 2009 NAS Report also recommended taking measures to attract students in the physical and life sciences to pursuing graduate studies in multidisciplinary fields critical to the practice of forensic science. Moreover, law students, law practitioners, and judges should be encouraged to acquire basic knowledge in these fields.⁴⁴²

The 2010 report of the NACDL, based on the 2009 NAS Report, offered more specific and detailed recommendations and added some new and crucial recommendations.⁴⁴³ These recommendations can be divided into seven major areas:

- (1) The establishment and funding of a central, science-based federal agency: The primary and central reform suggested is that Congress establish and allocate funds for a science-based federal agency, for the purpose of promoting “the development of forensic science into a field of multidisciplinary research and practice founded on the systematic creation, collection, and analysis of relevant data.” Validated and reliable forensic evidence is an important and vital component of the criminal justice system, and its development should be encouraged. Moreover, “[t]he results of any forensic theory or technique whose validity, limitations, and measures of uncertainty have not been established should not be admitted into evidence” in a criminal trial, and prior admissibility or use of the results of a forensic discipline, technique, or theory is not conclusive

441. *Id.* at 26 (Recommendation 9).

442. *Id.* at 27–28 (Recommendation 10).

443. NACDL-2010 Report, *supra* note 426, at 3–4.

proof of their validity or reliability. Accordingly, one of the agency's central and immediate priorities should be generating research programs for determining the validity, limitations, and measures of uncertainty of forensic theories or techniques, particularly with regard to forensic evidence that supposedly identifies any specific individual as being involved with a crime scene. This was followed by detailed recommendations regarding staffing, scope of responsibilities, an accreditation and certification board, and a proficiency testing program.⁴⁴⁴

- (2) Establishing a culture of science: The principle here is that “[a] culture of science that encourages independence, openness, objectivity, error management, and critical review should be promoted in forensic science practitioners and facilities.” This culture already exists among many forensic science practitioners and facilities, but a fundamental commitment to a culture of science should exist among *all* facilities and *all* practitioners. This was accompanied with detailed recommendations regarding autonomy, openness, objectivity, error management, and critical review.⁴⁴⁵
- (3) Setting a national code of ethics: The report expressed the principle that “[a]ll forensic science practitioners and supervisors should be required to adhere to a professional code of ethics that clearly articulates ethical obligations and contains a meaningful enforcement mechanism.” The detailed recommendations in this context relate to continuing education, acknowledgment of subjectivity, disclosure obligations, and enforcement.⁴⁴⁶
- (4) The institution of a prerequisite of research: The report recommended establishing and fully funding research programs relating to the accuracy, reliability, and validity of

444. *Id.*

445. *Id.* at 4–6.

446. *Id.* at 6–7.

forensic theories and techniques and their limitations and measures of uncertainty where calculable. This would be led and conducted principally by credentialed and qualified scientists at national research institutions with forensic science practitioners, particularly “those guided by a culture-of-science mindset and with histories of independence from law enforcement,” as active research participants and partners. Detailed recommendations were further given relating to determination of probability associations, relationships between research studies and case work, critical review, errors rates, automated techniques, bias minimization, and documentation.⁴⁴⁷

- (5) Improvement of education: The report noted that legal professionals generally lack the necessary scientific expertise and knowledge to understand and assess forensic evidence in an informed way. Thus, legal practitioners and judges must receive meaningful education and training “in the fundamentals of science, statistics, and common forensic practices; and in the limitations of, and potential forms and scope of error associated with, those practices.” The detailed recommendations related to law students, lawyers, and judges, as well as educational resources.⁴⁴⁸
- (6) The principle of transparency and disclosure: Transparency is vital to a fair and effective criminal justice system and a “hallmark of good science.” An attorney’s ability to evaluate, investigate, present, and confront forensic evidence at trial is contingent on complete and timely disclosure of information about the forensic examination, the conclusions of the forensic science practitioner, and the facility where the examination was conducted. “In every case involving forensic evidence, regardless of the current state of the science and/or advancements made, both the prosecution and the defense

447. *Id.* at 7–10.

448. *Id.* at 10–11.

will require full access to the forensic evidence and underlying data related to a particular case.” The detailed recommendations here related to transparency of forensic facility operations, ethical requirements, disclosure obligations, access to research and litigants, minimum disclosure requirements, reports, and databases.⁴⁴⁹

- (7) Allocating defense resources, particularly for indigent defense services: The principle articulated in this context is that “[f]orensic reform must be viewed within the framework of the fundamental constitutional protections established to ensure fair and accurate verdicts based on trustworthy evidence and to prevent wrongful convictions.” The prosecution tends to be the “primary proponent of forensic evidence,” but defense attorneys also sometimes use forensic evidence at trial. The report noted that many exonerations of innocently convicted defendants have been based on forensic evidence submitted by defense counsel. Defense counsel should, thus, be able to consult with forensic experts and experts in related scientific fields to present in court the scientific limits of the evidence, the results of independent testing, and the testimony of independent experts when appropriate. It was therefore recommended that the defense be ensured the necessary resources for obtaining such assistance from forensic and scientific experts and for the use of forensic facilities for independent, confidential testing. Indigent defendants, like defendants with financial means, moreover, should be ensured access to assistance from appropriate experts. The detailed recommendations addressed the topics of indigent defense, experts, consultation, and confidential testing.⁴⁵⁰

449. NACDL-2010 Report, *supra* note 426, at 11–14.

450. *Id.* at 14–16.

The last important report with recommendations is the 2016 Report to the President on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.⁴⁵¹

The study that led to the report was a response to [the President's] question to [the Council of Advisors], in 2015, whether there are additional steps on the scientific side, beyond those already taken by the Administration in the aftermath of the highly critical 2009 National Research Council report on the state of the forensic sciences, that could help ensure the validity of forensic evidence used in the Nation's legal system. [The Council] concluded that there are two important gaps: (1) the need for clarity about the scientific standards for the validity and reliability of forensic methods and (2) the need to evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable.⁴⁵² [The study] aimed to help close these gaps for a number of forensic "feature-comparison" methods—specifically, methods for comparing DNA samples, bitemarks, latent fingerprints, firearm marks, footwear, and hair.⁴⁵³

Here are some of the recommendations:

- (1) It is important that scientific evaluations of the foundational validity be conducted, on an ongoing basis, to assess the

451. EXEC. OFFICE OF THE PRESIDENT, *supra* note 130, at 1.

452. *Id.*

453. *Id.*

foundational validity of current and newly developed forensic feature-comparison technologies. To ensure the scientific judgments are unbiased and independent, such evaluations should be conducted by an agency which has no stake in the outcome.⁴⁵⁴

- (2) The National Institute of Standards and Technology (NIST) should take a leadership role in transforming three important feature-comparison methods that are currently subjective—latent fingerprint analysis, firearms analysis, and under some circumstances, DNA analysis of complex mixtures—into objective methods.⁴⁵⁵
- (3) The NIST should improve the Organization for Scientific Area Committees (OSAC), which was established to develop and promulgate standards and guidelines to improve best practices in the forensic science community.⁴⁵⁶
- (4) The Office of Science and Technology Policy (OSTP) should coordinate the creation of a national forensic science research and development strategy.⁴⁵⁷
- (5) The FBI Laboratory should undertake a vigorous research program to improve forensic science, building on its recent important work on latent fingerprint analysis.⁴⁵⁸
- (6) The Attorney General should direct attorneys appearing on behalf of the Department of Justice to ensure expert testimony in court about forensic feature-comparison methods meets the scientific standards for scientific validity.⁴⁵⁹
- (7) Where empirical studies and/or statistical models exist to shed light on the accuracy of a forensic feature comparison method, an examiner should provide quantitative information about error rates. In testimony, examiners should always state

454. *Id.* at 14.

455. *Id.* at 15.

456. EXEC. OFFICE OF THE PRESIDENT, *supra* note 130, at 15.

457. *Id.* at 16.

458. *Id.* at 17.

459. *Id.* at 18.

clearly that errors can and do occur, due both to similarities between features and to human mistakes in the laboratory.⁴⁶⁰

- (8) When deciding the admissibility of expert testimony, federal judges should consider the appropriate scientific criteria for assessing scientific validity, including “foundational validity” and “validity as applied.” Federal judges, when permitting an expert to testify about a foundationally-valid feature-comparison method, should ensure that testimony about the accuracy of the method and the probative value of proposed identifications is scientifically valid in that it is limited to what the empirical evidence supports. Statements suggesting or implying greater certainty are not scientifically valid and should not be permitted. In particular, courts should never permit scientifically indefensible claims such as: “zero,” “vanishingly small,” “essentially zero,” “negligible,” “minimal,” or “microscopic” error rates; “100 percent certainty” or proof “to a reasonable degree of scientific certainty;” identification “to the exclusion of all other sources;” or a chance of error so remote as to be a “practical impossibility.”⁴⁶¹

CONCLUSION

To the important and detailed recommendations made in the 2009 NAS Report, the 2012 NACDL Report, and the 2016 Report to the President, I would add three general recommendations that implement three fundamental safety rules, and a fourth, unique recommendation, which I raised here.

The first safety recommendation is that a legal rule must be set that precludes convicting on the basis of any single piece of evidence.⁴⁶² The rationale for this rule is that errors arise in every scientific test

460. *Id.* at 19.

461. *Id.*

462. We have seen the tremendous significance of the error rate. *See supra* Section II.C; *see also* Sangero & Halpert, *supra* note 59, at 43.

and that the possibility of an error, which is not negligible, prevents achieving proof of guilt beyond a reasonable doubt when based solely on a single piece of evidence. The second safety recommendation relates to regulation: manufacturers of forensic scientific equipment and forensic labs must be subject to safety regulation similar to the FDA's regulation of manufacturers of medical equipment and medical laboratories. The role of regulating scientific evidence must be shifted from judges to professional regulators with expertise in the relevant scientific fields,⁴⁶³ which is a standard practice in other areas with regard to safety-critical systems. The third safety recommendation is that those who will engage in safety in the legal field apply the error-prevention model developed and refined in the medical diagnostics field, as described above in section II.C. The fourth safety recommendation is the adoption of the MIT test I proposed above in discussing the distinction between scientific evidence and junk science. Under this selection-test, which would precede the Daubert examination of evidence presented as scientific, if there is no systematic scientific academic work in the relevant field, then the evidence is not scientific. Finally, further recommendations for improving forensic evidence should be developed by the proposed SCJSI or by the NIFS.

In line with the principles of safety, these recommendations must not, of course, be assumed to be exhaustive or necessarily well-suited to their goals. Rather, there is a need to revisit and check them after they have been implemented and to verify whether each one attains its objective, all in an attempt for unending improvement, as is the accepted practice in modern safety.

463. Halpert & Sangero, *supra* note 2, at 93–94.