

The Courts, the NAS, and the Future of Forensic Science

Jennifer L. Mnookin[†]

INTRODUCTION

On a recent flight, the person next to me on the crowded airplane began to chat with me. When I told her about what I researched and studied, she looked at me with a big grin. “I LOVE forensic science,” she said. “I watch CSI whenever I can. They can do such amazing things. It’s all so high tech—and incredibly accurate! It’s almost like magic, isn’t it?” She leaned in a bit closer and looked at me intently. “Tell me, is it like that in real life?”

I looked at her for a moment before answering. I felt a bit like the older child on the playground about to reveal to her younger friend that Santa Clause doesn’t really exist. I shook my head. “No, I wouldn’t say that CSI’s depiction is entirely realistic. In the real world, forensic science isn’t nearly so glossy. It isn’t nearly so speedy. And most important, it isn’t nearly so foolproof, either.”

“Really? That’s too bad,” she told me. She looked at me directly for a brief moment, shook her head, and then looked away. “Well, to tell you the truth, I think I’d rather just keep believing in the television version.” Figuring that reality was not going to be any match for CSI, I shrugged, and went back to the book I was reading.

In fact, that casual exchange on an airplane captures something quite important about the traditional forensic sciences, which find themselves at a crossroads. For many long-used types of forensic science, including fingerprint identification, firearms identification, handwriting

[†] Professor of Law, UCLA School of Law. This is an expanded and revised version of the Belfer Lecture, presented at Brooklyn Law School on April 7, 2009, in honor of Professor Margaret Berger’s retirement. Professor Mnookin thanks Margaret Berger, Ed Chang, Itiel Dror, Jennifer Friedman, Jay Koehler, and D. Michael Risinger for helpful comments, conversations and suggestions. Many thanks to Forrest Havens for his helpful research assistance.

identification, and toolmark identification, experts' claims about their field, the authority of their methodologies, and their own abilities have dramatically outstripped what has actually been established by persuasive research and careful study. Forensic scientists have regularly testified in court 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 the 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. Moreover, when academics attempt to do research on these questions, they have sometimes faced limited cooperation, or even downright resistance, from the forensic science community, because practitioners, managers, and laboratory directors (as well as police departments and prosecutors), are often wary of research not under their supervision or control.

For roughly the last decade, academic critics, and, occasionally, forensic scientists themselves, have argued that this state of affairs needs to change.¹ If we cannot trust the

¹ On forensic science in general, see Paul C. Giannelli, *Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs*, 86 N.C. L. REV. 163 (2007) [hereinafter Giannelli, *Wrongful Convictions*]; Roger Koppl, *How to Improve Forensic Science*, 20 EUROPEAN JOURNAL OF LAW AND ECONOMICS, 255 (2005); Michael J. Saks, *Merlin and Solomon: Lessons From the Law's Formative Encounters with Forensic Identification Science*, 49 HASTINGS L.J. 1069 (1998) [hereinafter Saks, *Merlin and Solomon*]; Michael J. Saks & David L. Faigman, *Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It*, 4 ANN. REV. L. & SOC. SCI. 149 (2008); Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCI. 892 (Aug. 5, 2005); Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199 (2008) [hereinafter Saks & Koehler, *Individualization Fallacy*]; Michael J. Saks & Jonathan J. Koehler, *What DNA 'Fingerprinting' Can Teach the Law About the Rest of Forensic Science*, 13 CARDOZO L. REV. 361 (1991). On fingerprint evidence, see e.g., SIMON A. COLE, SUSPECT IDENTITIES: A HISTORY OF FINGERPRINTING AND CRIMINAL IDENTIFICATION (2001) [hereinafter COLE, SUSPECT IDENTITIES]; David A. Stoney, *Fingerprint Identification: Scientific Status*, in MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 55 (David L. Faigman et al eds., 1997); David Stoney, *Measurement of Fingerprint Individuality*, in ADVANCES IN

evidentiary inputs into our criminal justice system, we cannot trust the outputs either. Numerous well-publicized wrongful convictions have made the danger of error in our criminal justice system both more obvious and more salient.² Recent research suggests that misleading and erroneous forensic science has been a significant contributing factor in many of the known wrongful convictions.³

FINGERPRINT TECHNOLOGY 327 (Henry C. Lee & R.E. Gaensslen eds., 2001); Christophe Champod, *Fingerprint Examination: Towards More Transparency*, 7 L. PROBABILITY & RISK 111 (2008); Simon A. Cole, *Does 'Yes' Really Mean Yes? The Attempt to Close Debate on the Admissibility of Fingerprint Testimony*, 45 JURIMETRICS J. 449 (2005); Simon A. Cole, *Fingerprinting: The First Junk Science?*, 28 OKLA. CITY U. L. REV. 73 (2003); Simon A. Cole, *Grandfathering Evidence: Fingerprint Admissibility Rulings from Jennings to Llera Plaza and Back Again*, 41 AM. CRIM. L. REV. 1189 (2004) [hereinafter Cole, *Grandfathering Evidence*]; Simon A. Cole, *The Prevalence and Potential Causes of Wrongful Conviction by Fingerprint Evidence*, 37 GOLDEN GATE U. L. REV. 39 (2006); Lyn Haber & Ralph N. Haber, *Scientific Validation of Fingerprint Evidence Under Daubert*, 7 L. PROBABILITY & RISK 87 (2007); Jonathan J. Koehler, *Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter*, 59 HASTINGS L.J. 1077 (2008) [hereinafter Koehler, *Proficiency Tests*]; Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 BROOK. L. REV. 13 (2001) [hereinafter Mnookin, *Fingerprint Evidence*]; Jennifer L. Mnookin, *The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate*, 7 L. PROBABILITY & RISK 127 (2008) [hereinafter Mnookin, *Confessions of a Fingerprinting Moderate*]; David A. Stoney & John I. Thornton, *A Critical Analysis of Quantitative Fingerprint Individuality Models*, 4 J. FORENSIC SCI. 1187 (1986) On handwriting identification, see Jennifer L. Mnookin, *Scripting Expertise: The History of Handwriting Identification Evidence and the Judicial Construction of Reliability*, 87 VA. L. REV. 1723 (2001) [hereinafter Mnookin, *Scripting Expertise*]; D. Michael Risinger, *Goodbye to All That, or A Fool's Errand, by One of the Fools: How I Stopped Worrying About Court Responses to Handwriting Identification (and 'Forensic Science' in General) and Learned to Love Misinterpretations of Kumho Tire v. Carmichael*, 43 TULSA L. REV. 447 (2007); D. Michael Risinger & Michael J. Saks, *Science and Nonscience in the Courts: Daubert Meets Handwriting Identification Expertise*, 82 IOWA L. REV. 21 (1996) [hereinafter Risinger & Saks, *Science and Nonscience in the Courts*]; D. Michael Risinger, Mark P. Denbeaux & Michael J. Saks, *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification 'Expertise'*, 137 U. PA. L. REV. 731 (1989). On ballistics evidence, see Adina Schwartz, *A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification*, 6 COLUM. SCI. & TECH. L. REV. 2 (2005).

² For an early account of the work of the Innocence Project, see BARRY SCHECK ET AL., *ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION, AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED* (2000); see also Brandon J. Garrett, *Judging Innocence*, 108 COLUM. L. REV. 55 (2008); D. Michael Risinger, *Innocents Convicted: An Empirically Justified Factual Wrongful Conviction Rate*, 97 J. CRIM. LAW & CRIMINOLOGY 761 (2007); The Innocence Project, <http://www.innocenceproject.org> (last visited Feb. 25, 2010).

³ Brandon J. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1 (2009). For analytic discussions of the role of forensic science in wrongful convictions, see, e.g., William C. Thompson, *Beyond Bad Apples: Analyzing the Role of Forensic Science in Wrongful Convictions*, 37 SW. U. L. REV. 1027 (2008); Jennifer L. Mnookin, *Excerpt from Transcript of Proceedings: Experts and Forensic Evidence*, 37 SW. U. L. REV. 1009 (2008); Jane Campbell Moriarty, "Miscorrections," *Science, and The Ministers of Justice*, 86 NEB. L. REV. 1 (2007).

While the danger of erroneous conviction provides both a moral and practical perspective on why reliable and valid forensic science is so important, ordinary expert-evidence doctrine also mandates its validity as a prerequisite for admissibility. In 1993, the Supreme Court, in *Daubert v. Merrell Dow*, made clear that judges have a gatekeeping responsibility with respect to expert evidence.⁴ In the federal courts and in those states that have embraced *Daubert*, expert evidence needs to be sufficiently reliable—meaning, more or less, scientifically valid—in order to be legitimately admissible in court.⁵ As a matter of formal evidence doctrine, then, forensic science evidence should only be permitted if it meets *Daubert*'s requirements. While *Daubert* envisioned the judicial gatekeeper's inquiry into reliability as "flexible," and therefore did not set up any absolute criteria for determining the validity (and hence admissibility) of expert testimony, the majority opinion did provide some important guidelines for trial court judges. Specifically, *Daubert* invites courts to look at whether the evidence or technique in question has been tested adequately, whether it has a known error rate; whether it has been subject to peer review; and whether it is generally accepted by the relevant scientific community.⁶ At this point, numerous *Daubert* challenges have been made to many kinds of forensic science, from fingerprint evidence, to ballistics analysis, to handwriting examination.⁷ But with a small

⁴ *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 597 (1993).

⁵ Justice Blackmun stated that the "inquiry" into the admissibility of expert evidence is "a flexible one," whose "overarching subject is the scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission." *Daubert*, 509 U.S. at 594-95.

⁶ *Id.* *Daubert*'s approach was expanded upon in *G.E. v. Joiner*, 522 U.S. 135 (1997) and *Kumho Tire v. Carmichael*, 526 U.S. 137 (1999).

⁷ For examples of courts permitting forensic science under *Daubert* or the state equivalent, see *United States v. Havvard*, 260 F.3d 597 (7th Cir. 2001) (finding that latent fingerprint identification methods satisfied the standards of reliability set forth by *Daubert*); *United States v. Mitchell*, 145 F.3d 572, 574, 579-80 (3d Cir. 1998); see also *United States v. Ford*, 481 F.3d 215, 222 (2d Cir. 2007) (finding shoeprint analysis to satisfy the *Daubert* standard); *United States v. Mitchell*, 365 F.3d 215, 233, 250 (3d Cir. 2004); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1180 (D.N.M. 2009) (finding firearm identification evidence to satisfy the *Daubert* standard); *State v. Foreman*, 954 A.2d 135 (Conn. 2008); *United States v. Mahone*, 453 F.3d 68, 79 (1st Cir. 2006); *United States v. Prime*, 363 F.3d 1028, 1034 (9th Cir. 2004) (finding that handwriting analysis satisfies the *Daubert* standard); *United States v. Crisp*, 324 F.3d 261, 271 (4th Cir. 2003); *Howard v. State*, 853 So. 2d 781, 796 (Miss. 2003) (finding that bite mark analysis satisfies the *Daubert* standard). For a list of *Daubert* challenges to fingerprint evidence (but current only through 2005), see http://onin.com/fp/daubert_links.html (last visited Feb. 26, 2010). For the argument that courts have been less intense in their *Daubert* scrutiny in criminal cases than in civil, see, e.g., D.

number of exceptions, courts have continued to permit these kinds of evidence without limit.⁸

The truth of the matter is that for the last decade, both judges and the forensic science community have chosen to behave rather like my acquaintance on the airplane. They prefer to play a kind of make-believe; they prefer to believe in the television version.

In what follows, I aim to do three things. First, in Part I, I will provide a brief overview of the present state of affairs within forensic science, focusing on latent fingerprint evidence and the concerns that have emerged regarding the adequacy of its research basis. Second, in Part II, I will discuss a recent and significant report issued in February 2009 by the National Academy of Sciences regarding the needs of the forensic science community.⁹ I will suggest that this report, though impressive in many ways, gave too little attention to how the courts ought to handle the admissibility of pattern identification evidence. I wholeheartedly agree with this report's assertion that it is imperative that we create substantially more funding for research and government oversight and regulation of forensic science. But this alone will not be enough, nor is it likely even to happen at all, unless courts also begin to take their responsibilities in this area more seriously. In Part III, I will look closely at how the courts have confronted (or, more accurately, mostly avoided confronting) the present problems relating to forensic science evidence in recent years. I will describe the approaches, mechanisms and machinations by which numerous courts have failed to treat their responsibilities to assess the validity of forensic science

Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on The Dock?*, 64 ALB. L. REV. 99 (2000).

⁸ There have been a small handful of cases that have restricted forensic science evidence, at least to some extent; this issue will be discussed in detail below, at *infra*, Part III (c). See, e.g., *Taylor*, 663 F. Supp. 2d at 1180; *United States v. Green*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005); *United States v. Hines*, 55 F. Supp. 2d 62, 73 (D. Mass. 2002); *United States v. Llera Plaza (I)*, 179 F. Supp. 2d 492 (E.D. Penn. 2002) (*overruled by Llera Plaza II*, 188 F. Supp. 2d 549 (E.D. Penn. 2002)); *Maryland v. Rose*, No. K06-0545 (MD Cir. Ct. Oct. 19, 2007). Expert evidence in handwriting identification has been scrutinized more carefully by courts than have the other forms of pattern identification evidence, and this scrutiny has sometimes led judges not merely to limit but to exclude it altogether. For a thoughtful account of the current case law in this area, see D. Michael Risinger, *Handwriting Identification*, in FAIGMAN, MODERN SCIENTIFIC EVIDENCE, *supra* note 1; D. Michael Risinger, *Cases Involving the Reliability of Handwriting Identification Since the Decision in Daubert*, 43 TULSA L. REV. 477 (2008) [hereinafter Risinger, *Cases Involving*].

⁹ STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (The National Academies Press 2009) [hereinafter NAS Report].

evidence with adequate care, and I will also describe the approaches taken by the few courts who have addressed the issues seriously. This section also offers the first detailed scholarly analysis of an approach taken by a handful of thoughtful jurists with respect to forensic pattern evidence, in which they continue to admit the evidence but only in a weakened, limited form—specifically, they permit the expert to describe similarities and differences between exemplars, while excluding the experts' ultimate conclusions about identification. This section describes why that approach, though superficially quite appealing, is far less conceptually coherent than it appears, though I recognize that it may nonetheless remain an attractive, pragmatic, stop-gap measure for courts wrestling with these difficult issues. Finally, in Part IV, I describe what a serious judicial examination of forensic pattern identification evidence ought to entail, what questions judges should focus on under *Daubert v. Merrell Dow*, and what demands they should make as a prerequisite to admissibility in court. I suggest that outright exclusion may, in some cases, indeed be warranted, and should certainly, along with more modest measures, be part of the available judicial toolkit.

A few brief preliminaries are needed to provide context and background. First, it is important to recognize that these concerns are not merely abstract or theoretical. In fact, a recent study of wrongful convictions found that flaws with forensic science—including interpretive errors, overstated testimony, and inaccuracies—were present in a whopping 60% of the cases studied.¹⁰ To be sure, many of the forensic science errors found

¹⁰ Garrett & Neufeld, *supra* note 3, at 19 n.62. In the case of microscopic hair analysis, it is important to distinguish those cases in which forensic scientists engaged in misstatement, overstatement, and unjustified assertions from those cases in which the inherently limited sensitivity of the technique means that in retrospect, the jury likely made an incorrect inference from the forensic evidence. Practitioners of microscopic hair identification have never formally claimed an ability to identify an individual from a hair; at most, a hair can be said to be “consistent” with a source, which means only that it is a member of a class of hairs that could possibly have come from that source, not that the hair necessarily came from that source. In this sense, it is more like blood typing—e.g., “this blood is type A, and so is the defendant, so the defendant is not excluded from the group of people who are possible sources for this blood” than it is like DNA profiling or fingerprinting. In my opinion, if a microscopic hair examiner provided only this kind of “class” evidence, including the defendant (or other relevant person) as a possible source among other possible sources, it ought *not* to be considered a forensic error, even if subsequent evidence shows the defendant *not* to be the perpetrator. It is as if the (innocent) defendant, the blood sample, and the actual perpetrator all had Type A blood—for the serologist to have testified that the defendant’s blood type matched the crime scene was not erroneous, though to be sure, the inference of guilt the jury drew from that fact combined with the other evidence

in the study related to techniques that play a less significant role as legal evidence than they once did, such as microscopic hair examination (which now most frequently functions as an adjunct to mitochondrial DNA testing of hairs instead of standing alone as it used to do).¹¹

However, to intimate that therefore these past errors and mischaracterizations no longer matter is to miss a critical point. While mitochondrial DNA testing can now often operate as a check on microscopic hair analysis, the broader forensic science ‘culture’ which made these earlier instances of error possible remains very much intact. This forensic science culture—a culture in which claims derived from experience are often accepted as a substitute for data; a culture in which interpretations are often framed in absolute terms rather than in more limited or modest language; a culture in which potentially biasing information is not systematically kept from the forensic examiner; and a culture in which institutionally cozy relationships between detectives, forensic analysts, and prosecutors may encourage unconscious partisanship—remains very much the norm within forensic science laboratories today. It is, in the end, this culture that needs to change; new and improved forensic techniques will not, by themselves, provide an adequate solution.

Second, even though I certainly do wish to criticize the current paltry research basis of forensic science and the courts’ response to that reality, I also want to emphasize several important caveats. Forensic professionals are, for the most part, just that—hard-working, dedicated, and trying their best, often with quite inadequate funding. Though there are, unfortunately, more than a handful of known instances of forensic fraud, these incidents are the exception and not the

was erroneous. By contrast, when, as was all too often the case, microscopic hair analysts claimed or intimated that the defendant was *the* source, or provided fictitious frequency estimates for hair types, or claimed the hairs “matched” without making clear that numerous other peoples’ hair in any given population would also match, this would, in my view, certainly count as forensic error. For examples of forensic error, see generally *id.* I thank Barry Scheck for useful conversations on this point that clarified my thinking, even if we continue to disagree on some definitional points.

¹¹ See, e.g., Cary T. Oien, *Forensic Hair Comparison, Background Information for Interpretation*, 11 FORENSIC SCI. COMM. 2009, available at http://www.fbi.gov/hq/lab/fsc/backissu/april2009/review/2009_04_review02.htm; SWGMAT (Scientific Working Group on Materials Analysis), *Forensic Human Hair Examination Guidelines*, 7 FORENSIC SCI. COMM. 2005, available at http://www.fbi.gov/hq/lab/fsc/backissu/april2005/standards/2005_04_standards02.htm#p11 (pointing out that microscopic hair analysis and DNA often complement one another).

rule.¹² Most forensic scientists are both diligent and honest. My criticisms of the field should not be heard as criticisms of all those who pursue it. Furthermore, the lack of an adequate research basis for the claims of the pattern identification sciences is *not* the fault of practicing rank-and-file forensic scientists themselves, most of whom have neither the training nor the background to pursue such research effectively even if they wanted to do so. These practitioners should—and indeed must—become, to some extent, the *subjects* of research inquiry, but they are not to be faulted for failing to become empirical researchers themselves.

Third and finally, even though I strongly believe that forensic science needs to be placed on a more secure research foundation, I do not want to suggest, or to be heard to suggest, that it is therefore of no value. In fact, with most of the forensic sciences, my strong suspicion is that when we do finally insist on pursuing the necessary research, we will find that many kinds of forensic evidence presently in use turn out to be extremely probative, and very much worth hearing in court. In many fields, my prediction would also be that the error rate, even with the methodologies in use at present, will turn out to be tolerably low for a wide array of pattern identification tasks, though I am far from confident this will turn out to be true in all fields, or in all situations. But for now, these expectations are mere speculation—the critical point is that it is time to pursue research that will help us find out for sure. What we have, at present, is no more and no less than an absence of adequate evidence. This lack of evidence does not in and of itself prove the inadequacy of the methods used by forensic scientists. But we can, should, and therefore must, seek better evidence establishing the validity of these techniques, methodologies, and conclusions, so that we can have greater legitimate confidence in the forensic sciences we use in court, and thus better understand their possible limits and weaknesses as well.

¹² See generally Keith A. Findley, *Innocents At Risk: Adversary Imbalance, Forensic Science, and the Search for Truth*, 38 SETON HALL L. REV. 893 (2008); Jennifer L. Mnookin, *Excerpt from Transcript of Proceedings: Experts and Forensic Evidence*, 47 SW. U. L. REV. 1009 (2008); Thompson, *supra* note 3; Paul C. Giannelli, *The Abuse of Scientific Evidence in Criminal Cases: The Need for Independent Crime Laboratories*, 4 VA. J. SOC. POL'Y. & L. 439 (1997); Giannelli, *Wrongful Convictions*, *supra* note 1; Moriarty, *supra* note 3. For examples of forensic fraud, see, e.g., Forensic Fraud, http://www.corpus-delicti.com/forensic_fraud.html (last visited Apr. 13, 2010) (detailing many such cases).

I. FINGERPRINT EVIDENCE AND THE LIMITS OF OUR KNOWLEDGE

In what follows, I focus primarily on latent fingerprint identification, but it is important to realize that I could tell an extremely similar tale about a variety of other kinds of forensic science, including firearms identification, handwriting identification, bitemark identification, toolmark identification, and the like. 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. It therefore provides an especially good focus, for if the problems I am describing exist within *this* forensic domain, they are likely to be equally or more acute in other areas of pattern identification.

Fingerprint evidence was first used in the American courtroom nearly a century ago in 1911, and for most of its history it has been seen as the “gold standard” of forensic science. In recent years, however, whatever metal out of which this evidentiary standard was made has rather noticeably begun to tarnish.¹³

The basic approach taken by latent fingerprint experts involves what they call ACE-V. This acronym stands for analysis, comparison, evaluation and verification.¹⁴ First, in the analysis step, the examiner looks closely at the latent print associated with the crime at issue, and decides whether there is enough useful information contained in the image that it is “of value” for further examination.¹⁵ If so, the examiner then looks carefully at the various minutiae that he or she sees in the image, and, depending on local practices and the apparent difficulty of the print, typically marks up the print and documents the minutiae she observes.¹⁶ Second, in the comparison stage, the expert compares the latent print to a

¹³ For the history of fingerprint identification and its legal use, see generally COLIN BEAVEN, *FINGERPRINTS: THE ORIGINS OF CRIME DETECTION AND THE MURDER CASE THAT LAUNCHED FORENSIC SCIENCE* (2001); COLE, *SUSPECT IDENTITIES*, *supra* note 1; Simon A. Cole, *Witnessing Identification: Latent Fingerprint Evidence and Expert Knowledge*, 28 SOC. STUD. SCI. 687 (1998); Mnookin, *Fingerprint Evidence*, *supra* note 1.

¹⁴ For a description of each stage, see, e.g., Michell Triplett's Fingerprint Dictionary, available at <http://www.nwlean.net/fprints/a.htm>.

¹⁵ *Id.*

¹⁶ See, e.g., Herman Bergman & Arie Zeelenberg, *Fingerprint Matching*, Manual, in *Encyclopedia of Biometrics* (Anil K. Jain & Stan Z. Li., eds), at 502-04 (2009).

particular source print, noting both observed similarities and differences.¹⁷ Third, the analyst evaluates these similarities and differences, and reaches one of three, and only three, conclusions: identification, exclusion, or inconclusive.¹⁸ Note that these are the *only* permissible options available to a latent fingerprint expert—a match, a non-match, or a conclusion of “I don’t know.” “Maybe,” “possibly,” and “probably,” are not determinations presently permitted to fingerprint examiners under their professional rules and norms.¹⁹ Finally, in the verification step, if the first examiner has determined that the prints match, a second examiner takes the prints and goes through the same process to re-analyze them. In most laboratories, this step is conducted by an examiner who is informed of the original examiner’s conclusion before undertaking his or her own analysis.²⁰ This verifying examiner typically recognizes both that (a) he or she is verifying a conclusion already reached by someone else; and (b) that the conclusion already reached is that the prints do match.

Latent fingerprint examiners regularly claim that ACE-V is a version of the scientific method and assert that it offers a reliable methodology that establishes that fingerprint evidence is indeed a valid science.²¹ Many courts have agreed that ACE-V passes muster under *Daubert*.²²

¹⁷ The comparison print may have been to a known suspect, or to a non-suspect known to have been in the relevant location. Or the latent print may be submitted for an Automated Fingerprint Identification System (AFIS) database search, which compares the latent to a large database (depending on the particular database, often many millions of images) and returns a set of possible prints for human comparison. For some of the difficulties with interpretation of match thresholds in AFIS searches, see Itiel Dror & Jennifer Mnookin, *The Use of Technology in Human Expert Domains: Challenges and Risks Arising from the Use of Automated Fingerprint Identification Systems in Forensic Science*, 10 L. PROBABILITY. & RISK (forthcoming, 2010).

¹⁸ For descriptions of these permitted conclusions, see SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study, and Technology), Standard for Conclusions, http://www.swgfast.org/Standards_for_conclusions_ver_1.0.pdf.

¹⁹ Note that at its 2009 annual meeting, the International Association of Identification (IAI), the professional organization of fingerprint examiners, considered eliminating the longstanding professional restriction on testifying in probabilistic terms. However, the motion was the subject of significant contention and was tabled for further study.

²⁰ The FBI has begun to conduct some verifications “blind,” meaning that the verifier does not recognize that he or she is verifying a conclusion reached by another examiner. See Robert B. Stacey, *A Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 54 J. FORENSIC IDENTIFICATION 706, 715 (2004).

²¹ See, e.g., Peter E. Peterson et al., *Latent Prints: A Perspective on the State of the Science*, 11 FORENSIC SCI. COMM. 2009, available at http://www.fbi.gov/hq/lab/fsc/current/review/2009_10_review01.htm. Numerous fingerprint examiners have

In fact, ACE-V's relationship to the scientific method is tenuous at best: as a methodology, it amounts, more or less, to having *two different examiners look carefully at a set of fingerprints*. To be sure, the "scientific method" is itself a complicated and capacious idea, not altogether easily or adequately defined. But however we might define the critical characteristics of the scientific method, it surely amounts to more than simply careful, semi-structured observation. At root, ACE-V in its current incarnation amounts to no more and no less than a set of procedures to describe the careful comparison of a latent print with a potential source print by an initial examiner and a subsequent verifier. While careful observation and the recording of one's observations may be a necessary *part* of many scientific practices, careful observation in and of itself cannot be meaningfully said to constitute a *method*. Moreover, the simple act of labeling this process of careful observation as a methodology does not make it into one. Nor does bestowing upon it the label "scientific" tell us, through the moniker, anything about its likely validity or error rate.²³

The basic difficulty is that ACE-V is too general in conception and scope to provide much in the way of guidance or constraint for those who practice it. The devil is in the details—what constitutes analysis? How exactly does a competent comparison take place? When are apparent similarities misleading, and when might apparent differences be attributed to something other than the two prints deriving from different sources? ACE-V, as a methodology, does not help answer any of these critical methodological questions, because its categories are too general and insufficiently substantive.²⁴

It is as if one were to describe the methodology for fixing a car by the acronym DACT—Diagnose, Acquire, Conduct, and Test. We could describe the DACT car-repair methodology as

described ACE-V as corresponding to the scientific method. See e.g., Mary Beeton, *Friction Ridge Identification Process—Proposed Scientific Methodology*, THE DETAIL, Feb. 18, 2002, available at <http://www.clpex.com/Articles/TheDetail/1-99/TheDetail28.htm>.

²² See, e.g., United States v. Pena, 586 F.3d 105 (1st Cir. 2009); United States v. Vargas, 471 F.3d 255 (1st Cir. 2006); Crisp, 324 F.3d 261; United States v. Sullivan, 246 F.Supp.2d 700 (E.D. Ky. 2003); Markham v. State, 984 A.2d 262 (Md. App. 2009); Burnett v. State, 815 N.E.2d 201 (Ind. App. 2004).

²³ For the argument that the point ought not to be whether fingerprint evidence is or is not scientific, but rather, how to improve it, see Itiel E. Dror, *How Can Francis Bacon Help Forensic Science? The Four Idols of Human Biases*, 50 JURIMETRICS J. 93 (2009).

²⁴ See generally Haber & Haber, *supra* note 1.

follows: (1) *diagnose* the car's problem, (2) *acquire* the necessary parts for the repair, (3) *conduct* the repair, and (4) *test* to verify that the repair fixed the problem. Whether or not such a car-repair methodology actually works, or how well it works, would depend entirely on the content given to these very broad categories in specific instances. If in fact, someone diagnosed the car's problem correctly, located the appropriate parts, and conducted the repair properly, the methodology would work. But if the mechanic misdiagnosed the difficulty, acquired the wrong parts, or made an error when conducting the repair, the repair would fail, even though he or she had, in some sense, followed the methodology. Now, in light of the failed repair effort, a defender of ACE-V (and DACT) might suggest that the mechanic had not in fact followed DACT correctly, because he or she misdiagnosed the problem, made an error in the repair, or made some other mistake in application. The DACT defender might even argue that the mechanic's failure to fix the car *established* that she failed to follow DACT; that following DACT necessitates doing the steps *correctly*, not just *endeavoring* to follow them. But that response would render DACT (or, analogously, ACE-V), in some sense, merely tautological. The method does not describe with any specificity how to complete its requirements correctly. It is therefore illegitimate to argue that the method has not been followed simply because the desired outcome did not occur, precisely because the method itself underspecifies what is required. DACT itself does not explain how to diagnose, or what constitutes a sufficient repair, just as ACE-V does not explain how to analyze or compare (beyond calling for careful looking at a target portion of each print), or what constitutes a sufficient evaluation.²⁵ To be sure, ACE-V might be a useful description of the basic steps a fingerprint examiner takes in order to conduct his or her examination, but that does not make it a very useful description of a methodology, much less a so-called "scientific method."

²⁵ I am presently a member of an NIST/NIJ working group on Human Factors in Latent Fingerprint Examination. As part of this working group's efforts, fingerprint examiners put together a process map describing the ACE-V process in more careful detail than had ever previously occurred. While the process map is useful in many ways, it also makes my point here clear: although the process map delineates numerous steps, and diagrams in detail the steps necessary for conducting a comparison, the actual interpretive content of terms like "sufficiency," "identification," etc., are never actually specified, because at present, the fingerprint community does not have shared definitions of these concepts.

Surely, one would think, ACE-V in practice must amount to more than I am suggesting? Latent print examiners do have norms about what kinds of print ridge detail and minutiae they ought to be looking at, and examiners are trained to search both for relevant minutiae and to assess their contextual relationship and position on a fingerprint.²⁶ And latent print examiners do discuss with one another, informally, their personal notions regarding sufficiency, or the virtues and limitations of different categories of print information.²⁷

While individual examiners or even sometimes laboratories may develop working rules of thumb about the quantity of similarity required, latent 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.²⁸ Instead, examiners decide for themselves, based on their training and experience, how much similarity is sufficient to declare a match. Moreover, when examiners look at a print, they may not even be focusing on the same features. Two fingerprint analysts will often focus on different minutiae in their examination of the same print; indeed, sometimes the same examiner, when given the same print at a different time, will focus on different minutiae than

²⁶ Level 1 detail describes the major pattern of the print, whether it is, say, a tented arch or a loop. Level 2 detail, which constitutes the main focus for comparison and evaluation for most examiners, refers to the ridge quantities, details, and characteristics, such as whether and where a friction ridge bifurcates, or terminates, or develops a spur. Level 3 detail refers to sub-ridge detail, for example sweat pores that may be visible in an image.

²⁷ Fingerprint examiners, though, do not all agree that it is useful to divide the field into precisely three different levels. *See supra* note 25. Though the terms are often used in the United States, examiners in the UK do not typically focus on these categories, and there are diverging opinions on how to define and how to make use of Level 3 detail, for example. See the definitions of Level 1, 2, and 3 detail in Michelle Triplett's *Fingerprint Dictionary*, *supra* note 14 (available at <http://www.nwlean.net/fprints/1.htm>); *see generally* DAVID R. ASHBAUGH, *QUANTITATIVE-QUALITATIVE FRICTION RIDGE ANALYSIS: AN INTRODUCTION TO BASIC AND ADVANCED RIDGEOLOGY* (1999).

²⁸ Indeed, the IAI (International Association of Identification) resolved in 1973 that there was no scientific justification for having a specified minimum number of points of similarity in order to declare that two prints matched. *See Report of the Standardization Committee of the International Ass'n for Identification*, IDENTIFICATION NEWS, Aug. 1, 1973, available at <http://www.latent-prints.com/images/IAI%201973%20Resolution.pdf>. For further discussions of this point, see generally Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1; Simon A. Cole, *What Counts for Identity? The Historical Origins of the Methodology of Latent Fingerprint Identification*, 12 *SCI. IN CONTEXT* 139 (1999) [hereinafter Cole, *What Counts for Identity?*].

he or she did the first time.²⁹ The judgment is fundamentally a subjective one, not based on any formalized measures of either quantity or sufficiency.³⁰

Additionally, latent fingerprint examiners do not generally employ any statistical information or models in the ordinary ACE-V process. The field presently does not have or make use of robust statistically-based data about the frequency of different friction ridge characteristics. Analysts do not make regular or structured use in their comparisons of empirical studies showing how common or how rare different fingerprint details might be. They do not presently make use of any statistically-validated standards to justify how many identifying characteristics must be the same on two prints in order to warrant a finding that they match. Nor do they employ a probabilistic approach to determining the likelihood that a print selected at random would have that quantum of similarity, akin to the use of “random-match probability” in DNA identification. Although significant strides are being made toward developing these kinds of information, technical obstacles still limit the ability to develop a satisfactory statistical measure of the frequency of various ridge characteristics.³¹ As of now, there simply is no well-accepted, fully-specified statistical model that is available for latent fingerprint examiners to employ.

The list of difficulties continues. A fundamental tenet of latent fingerprint analysis is the “one discrepancy rule”—if there is even one genuine discrepancy between the latent print and a potential source print, then the two prints cannot have come from the same source.³² This, however, invites the critical question of how to decide what constitutes a discrepancy, as

²⁹ Itiel E. Dror et al., *Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Consistency and the Effect of a ‘Target’ Comparison* (unpublished article, under review at FORENSIC SCI. INT’L) (on file with author).

³⁰ See generally *id.*; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1; Mnookin, *Fingerprint Evidence*, *supra* note 1; Saks, *Merlin and Solomon*, *supra* note 1; Cole, *What Counts for Identity?*, *supra* note 28.

³¹ One quite promising approach is being developed by a group of European researchers. See, e.g., Cedric Neumann et al., *Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiae*, 52 J. OF FORENSIC SCI. 54, (2007); J.S. Buckleton, C.M. Triggs & C. Champod, *An Extended Likelihood Ratio Framework for Interpreting Evidence*, 46 SCI. AND JUST. 69 (2006). While a version of software implementing this team’s approach to providing probabilistic likelihood ratios is likely to be available soon, its adequacy and validity is not yet fully established.

³² OFFICE OF THE INSPECTOR GENERAL REPORT ON MAYFIELD (2006), <http://www.justice.gov/oig/special/s0601/final.pdf>; Triplett’s Fingerprint Dictionary, *supra* note 14.

opposed to a dissimilarity that can legitimately be explained in some other way. The problem is that *no* two print impressions are ever truly *identical*—every single impression from a print is distinct from every other impression of a print, different to some extent even from those that came from the same source. A print image can be affected by the pressure with which it was left, the surface on which it was made, the processes by which it was lifted, and many other factors. The question when comparing two prints, then, is not whether they are truly “identical”—for they will never be truly identical—but rather, whether they are sufficiently similar to each another to permit the conclusion that they came from the same source. The examiner needs to determine whether apparent differences are true dissimilarities, or instead, merely artifacts that ought not to be deemed meaningful. Unfortunately, latent fingerprint examiners lack any formalized criteria for determining when a difference between two prints is genuinely a dissimilarity, or when it might appropriately be explained in other ways. At root, this is again a matter of subjective judgment by the trained examiner.

Note, however, that the fact that these judgments are subjective does not necessarily imply that they are incorrect or unreliable. If I were to look at many different photographs of my sister, no two images of her would be identical. And yet, my judgment of whether any given photograph was an image of my sister or actually an image of someone else bearing a certain degree of resemblance to my sister would, I would wager, have a high probability of being correct. I would posit that my ability to identify images containing my sister, and to avoid misidentifying images of other people as my sister would be quite high—notwithstanding my lack of formal criteria for doing so. The absence of formal, validated standards for making such identification of my sister does not mean that I lack all relevant knowledge. My experience of many years of seeing my sister in a great variety of contexts would indeed likely help me with the identification tasks.

However, I also suspect that my ability to identify my sister in photographs would be strong but not perfect. In some images, she might be too far away, or too blurry, or someone else might bear such a strong resemblance to her, that despite my life-long knowledge of her from every angle, I might nonetheless mistake the other person for my sister. Or conversely, I might fail to recognize that some picture truly did show an image of my sister. Of course, to determine how often I

was right or wrong, we would also want to make sure we had a good method by which to determine “ground truth,” whether or not the photograph truly was of my sister.

The purpose of this analogy is to suggest three points that apply as much to fingerprint identifications as to my hypothetical efforts to identify my sister. First, I want to suggest that experience *can* be a legitimate basis for knowledge. Second, I want to suggest that knowledge need not necessarily be formalized to count as legitimate or valid. However, and this is the third point, if we wanted to find out just how good my ability to recognize my sister in photographs really was, we would need to depend on something that went beyond my say-so. We would not want simply to take my word for it when I said I was good at the task. We would not want to take the simple fact of my extensive experience looking at my sister as proof of my identification talents. Nor would we want to blindly accept my opinion that particular photos actually were or were not of my sister.

Instead of taking my say-so, my experience, and my conclusions as proof of my accuracy, we should carefully test my actual proficiency at the tasks. We would need to investigate empirically just how well I did identify my sister; in what array of circumstances I succeeded; and when and how often the task proved beyond my capacity. Indeed, as I will argue below, an equivalent focus on serious, careful proficiency testing of practitioners is precisely what we ought to demand in the realm of forensic science as well. Just as we would want proficiency testing to verify my claimed experience-based ability to identify my sister, so we also ought to require significant proficiency of fingerprint examiners and other pattern identification analysts.³³ And just as we ought not to simply take my assertions about my conclusions’ accuracy as proof of actual accuracy, we ought not to take fingerprint examiners’ experience-based assertions of accuracy as proof of accuracy either.³⁴

³³ See *infra* note 157 and accompanying text (discussing proficiency tests for fingerprinting).

³⁴ One important issue is feedback: if I do make mistakes in identifying my sister, am I likely to know about them? Do I receive feedback on my accuracy that could permit me to learn from my errors? If not, we should be especially skeptical of my assertions of accuracy based on experience, for if my experience is not likely to provide me with information about when and where I went wrong, I may not be able to learn all that much from it. This is clearly a concern with the pattern identification sciences as well, in which the opportunities for learning from feedback are quite limited. (The situation may be even worse than one in which there is no feedback; the trial process

Returning now to latent fingerprint examination, there are two additional difficulties with the current state of knowledge and practice. First, fingerprint experts claim to be able to *individualize*—to connect a given print to a *unique* source. When they declare a match, they assert that two prints come from a common source *to the exclusion of all other possible sources in the world*. This is an astonishingly strong claim.³⁵ A latent fingerprint examiner who individualizes is saying that he or she can connect this print to one particular finger of one single person, out of everyone in the world, everyone who has ever lived or will ever live. But there is quite simply a lack of empirical evidence establishing that they can actually do what they claim.³⁶ To be sure, both experience and some empirical research does suggest that fingerprints are highly varied.³⁷ But even if fingerprints themselves are unique, this does not necessarily mean that experts can make unique identifications from partial latent prints, using their methods and expertise. Evidence of uniqueness does not itself directly support the experts' claims that they are able to individualize.³⁸

may provide erroneous or misleading feedback in the case of forensic errors, precisely because forensic science evidence is often considered to be so strong. For example, if an expert makes an erroneous fingerprint identification that leads to a conviction or a guilty plea, that could be seen by the expert as confirming the correctness of the call, when in fact the erroneous identification may have substantially produced the conviction.)

³⁵ Champod, *supra* note 1. Saks & Koehler, *Individualization Fallacy*, *supra* note 1; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1, at 139-40; Simon A. Cole, *More than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985, 992-93 (2005) [hereinafter Cole, *More than Zero*]. See generally Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 L. PROBABILITY & RISK 233 (2009) [hereinafter Cole, *Forensics Without Uniqueness*].

³⁶ NAS Report, *supra* note 9; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1; David A. Stoney, *What Made Us Ever Think We Could Individualize Using Statistics?*, 31 J. FORENSIC SCI. SOC'Y 197, 197 (1991); Saks & Koehler, *Individualization Fallacy*, *supra* note 1.

³⁷ See, e.g., an unpublished study cited and described in *United States v. Mitchell*, 365 F.3d 215 (3d Cir. 2004) as the 50k study, designed to show the tremendous variation in fingerprints. But for a strong critique of this study, see David H. Kaye, *Questioning a Courtroom Proof of the Uniqueness of Fingerprints*, 71 INT'L STAT. REV. 521 (2003).

³⁸ See generally Cole, *Grandfathering Evidence*, *supra* note 1; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1. Think again of my sister in the photograph, discussed *supra* notes 33-34 and accompanying text. That every person really is unique does not mean that every photographic image will be a sufficiently clear depiction to permit a unique identification, or that my sister-identification talents will be strong enough to succeed in all circumstances. Even if every person (or every fingerprint) is indeed unique, the question is whether the image and the methods used to analyze it are capable of *discerning* that degree of difference.

Nonetheless, experts claim this ability as well as frequently asserting 100% certainty in their own conclusions. In fact, they are *prohibited* under their professional norms from making probabilistic judgments, and are subject to possible sanction if they do not follow this professional rule.³⁹ They are told and taught that they must either be absolutely certain, or reach no conclusion at all. There are no shades of grey permitted—notwithstanding that fundamentally, fingerprint matching ought to be thought of as a probabilistic inquiry.⁴⁰

How often do fingerprint examiners make mistakes? Well, in court, until quite recently, experts frequently testified that their technique had a “zero error rate.”⁴¹ Some examiners tried to divide their analysis of error rate into two parts—the error rate of the technique itself, and the error rate of the humans who use it. They acknowledged that it was possible for a human to make a mistake, though they asserted (again without any significant published research) that errors are exceedingly rare. They claimed that the technique, if used properly, is perfect and error-free. Mistakes result only when humans misapply it.⁴²

This notion of an error rate of zero is exceedingly unscientific. It borders on the meaningless, and is a far cry from how scientists typically think about error rates. Nothing is truly perfect—no human endeavor has an error rate of zero. Moreover, the distinction between the error rate of the technique and the error rate of the humans who use it is, frankly, nonsensical with regard to fingerprint identification. The human beings engaging in ACE-V *are* the technique. The appropriate question is the error rate in *practice*, not an in-the-clouds theoretical error rate that postulates perfect human

³⁹ See generally SWGFAST Standards for Conclusion, *supra* note 18; *supra* text accompanying note 19.

⁴⁰ Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1; Cole, *More than Zero*, *supra* note 35, at 992; Cole, *Forensics Without Uniqueness*, *supra* note 35, at 235. However, there is ongoing discussion in the fingerprint community about whether to modify or eliminate this rule. See *supra* text accompanying note 19; Stacey, *supra* note 20, at 715.

⁴¹ The FBI's Steven Meagher was perhaps the best known example of an expert who regularly testified in this vein. See, e.g., *Mitchell*, 365 F.3d at 222-26 (discussing Meagher's testimony); see also, e.g., *United States v. Mahone*, 453 F.3d 68 (Me. 2006) (zero error rate argued in footwear identification case, drawing on latent fingerprint for support and by analogy); *Havvard*, 260 F.3d 597; *Llera Plaza (I)*, 179 F. Supp. 2d 492.

⁴² See, e.g., *Mitchell*, 365 F.3d at 222-26 (discussing Meagher's testimony); *Llera Plaza (I)*, 179 F. Supp. 2d 492; see also Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1; Cole, *More than Zero*, *supra* note 35, at 1050.

beings and then concludes that so long as these perfect human beings make no mistakes, the error rate is zero. We could just as easily say that in theory, eyewitness identification has an error rate of zero because faces are in fact different— notwithstanding the fact that in practice, eyewitness identification errors are distressingly common.⁴³ This claim of an error rate of zero is an example of how the rhetoric of forensic sciences is often shaped for courtroom use rather than derived from valid scientific testing designed to produce the most accurate possible information.⁴⁴

If the actual error rate is not zero, then how often do fingerprint examiners make mistakes? The truth is that we really do not know. Although fingerprint examiners may, at times, undergo proficiency tests, these exams have for the most part been extremely easy, far easier than the kinds of challenges that can be faced in actual casework.⁴⁵ A fingerprint examiner from Scotland Yard once, under oath in court, referred to the FBI's proficiency tests as a "joke" because of how ridiculously easy they were.⁴⁶ Furthermore, in some laboratories, examiners take their proficiency tests in groups rather than individually. In addition, proficiency tests are generally not conducted blind, as part of what appears to the examiner to be ordinary casework.⁴⁷ Rather, examiners are usually aware they are being tested, and may therefore, consciously or unconsciously use a different degree of care than usual.⁴⁸ Thus, in their current form, proficiency tests might be a check on gross individual incompetence, but they certainly provide nothing more, and given that the examinations may be done collectively and not blind, they may not even necessarily provide that.

We do know that errors sometimes occur, though it is impossible on the basis of what we presently know to attempt

⁴³ See, e.g., BRIAN L. CUTLER & STEVEN D. PENROD, *MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY, AND THE LAW* (1995); ELIZABETH LOFTUS, *EYEWITNESS TESTIMONY* (1979).

⁴⁴ See Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1, at 137; see also Mnookin, *Scripting Expertise*, *supra* note 1, at 1727.

⁴⁵ See Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1, at 135-36; Koehler, *Proficiency Tests*, *supra* note 1, at 1092; Cole, *More than Zero*, *supra* note 35, at 1032.

⁴⁶ *Llera Plaza (II)*, 188 F. Supp. 2d 549.

⁴⁷ Koehler, *Proficiency Tests*, *supra* note 1, at 1092; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1, at 136.

⁴⁸ Koehler, *Proficiency Tests*, *supra* note 1, at 1092.

to quantify their frequency.⁴⁹ One particular fingerprint error—perhaps the monster of all fingerprint errors, the most high-profile, embarrassing fingerprint mistake in recent history, at least here in the United States—has contributed to shaping and framing the discourse surrounding latent fingerprint identification.⁵⁰ This mistake was sufficiently public, serious, and embarrassing that it led to a substantial inquiry into its causes; more generally, it made the fingerprint community—and the legal community—recognize that fingerprint errors were not simply a matter of incompetence or an issue of purely academic concern.⁵¹ I am referring, of course, to the mistaken identification of Brandon Mayfield, an attorney from Portland, Oregon who was held as a material witness in relation to the 2004 Madrid train bombing. The only evidentiary basis for suspecting his involvement was an alleged fingerprint match. Mayfield's print had been one of the possible source prints suggested by a computer database search using an AFIS (Automated Fingerprint Identification System). Mayfield's print appeared fourth down on the computer-generated list of suggestions—and according to the FBI, his print was a definite match.⁵²

Mayfield insisted that the identification had to be a mistake. He told authorities he had never set foot in Spain, had remained entirely in the United States during the relevant period, and indeed, lacked a passport. But three separate fingerprint examiners at the FBI, including two of the most respected senior examiners in the office, all concluded that the match was 100% certain. Even an independent, court-appointed expert confirmed the match as well.⁵³

⁴⁹ For an effort to describe the array of known errors, see Cole, *More than Zero*, *supra* note 35.

⁵⁰ The ongoing saga in the United Kingdom relating to Shirlie McKie might offer the Mayfield case some competition.

⁵¹ This increased awareness, openness, and a certain increased willingness to confront the limitations of the field, may indeed be the “silver lining” of the Mayfield case. See Jennifer Mnookin, Op-Ed., *The Achilles' Heel of Fingerprints*, WASH. POST, May 29, 2004, at A27.

⁵² Editorial, *The F.B.I. Messes Up*, N.Y. TIMES, May 26, 2004 at A22.; see generally OFFICE OF THE INSPECTOR GENERAL REPORT ON MAYFIELD (2006), <http://www.justice.gov/oig/special/s0601/final.pdf> [hereinafter OIG REPORT].

⁵³ Les Zaitz, *Transcripts Detail Objections, Early Signs of Flaws*, THE OREGONIAN, May 26, 2004, at A1; Andrew Kramer, *Fingerprint Science Not Exact, Experts Say*, ASSOCIATED PRESS, May 21, 2004, available at <http://www.msnbc.msn.com/id/5032168>; see also Steven T. Wax & Christopher J. Schatz, *A Multitude of Errors: The Brandon Mayfield Case*, 28 THE CHAMPION, Sept.-Oct. 2004, at 6; OFFICE OF THE INSPECTOR GENERAL REPORT ON MAYFIELD (2006), <http://www.justice.gov/oig/special/s0601/final.pdf>; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra*

The Spanish authorities were less convinced, and after several weeks, located another suspect, an Algerian named Ouhmane Daoud, in a different database, who, they claimed, was the actual source of the print. Eventually, the FBI concurred. The FBI was deeply embarrassed, Mayfield was released from custody, and eventually received compensation of \$2 million.⁵⁴

What happened? I will mention just two of the most important causes of the error. First, one portion of one of Brandon Mayfield's prints really does bear a striking resemblance to one portion of one of Ouhmane Daoud's fingers.⁵⁵ There is no doubt that portions of the two prints are extremely similar, and the resemblance between Mayfield's finger and the portion of the image most clearly visible on the latent recovered from Madrid was, as it happens, particularly strong. How often are we likely to see such a high degree of resemblance in prints from different sources? No one really knows. The Inspector General's report, an independent investigation conducted in the aftermath of the scandal, insists that this degree of similarity is extraordinarily rare.⁵⁶ Perhaps so, but the truth is that we do not actually know how common or rare that degree of apparent similarity may be. It is clear that the growing size of the databases used for fingerprint analysis increase the risks of misidentifications like this one.⁵⁷ Latent fingerprint examiners, at present, do make regular use of AFIS systems, computerized databases to generate a set of possible matching prints—possible “cold hits” based on a fingerprint match. But the computer algorithms are far from perfect, and thus the computer search process alone cannot determine whether any of the possible prints actually match. Only the examiner, using ACE-V to compare each AFIS suggestion to the latent print, can make that determination. As the databases grow, so grows the possibility of highly similar near-misses like Brandon Mayfield's—fingerprints so similar that they might fool even crack fingerprint experts.⁵⁸

note 1; Cole, *More than Zero*, *supra* note 35; Robert B. Stacey, A Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case. 54 J. Forensic Identification 706 (2004).

⁵⁴ Eric Lichtblau, *U.S. Will Pay \$2 Million to Lawyer Wrongly Jailed*, N.Y. TIMES, Nov. 30, 2006, at A18.

⁵⁵ OIG REPORT, *supra* note 52 at 6-7.

⁵⁶ *Id.* at 7.

⁵⁷ Dror & Mnookin, *supra* note 17, at 55-56.

⁵⁸ *Id.*

In addition to the unexpected degree of similarity between the prints from different sources, it appears that cognitive bias also played a role in the debacle. Immigration lawyer Mayfield was a Muslim; he had converted to Islam some time earlier.⁵⁹ He had also once represented a known terrorist in a child custody dispute. While it appears that the FBI investigators did not know these facts about Mayfield when they first determined that his print matched the one found on the detonation materials in Madrid, their subsequent awareness of this information made them more reluctant to reopen the issue or contemplate the possibility that they had made an error. More generally, even apart from this contextual information, it seems that once the first FBI examiner declared the prints to match, the verifying examiners *expected* to find a match. It is no great surprise, then, that they found precisely what they expected to find, likely the result of a mixture of peer pressure and expectation bias.⁶⁰

This problem of biasing information goes well beyond the Mayfield debacle. Forensic experts frequently have access to information about a case that goes beyond whatever information is actually necessary for their forensic testing. They may be told by detectives or investigators about other powerful evidence linking the suspect to the crime. They may be told details about the suspect that bear no relation to the pattern identification evidence itself—that he is a known gang member, or that she has prior convictions, or that he has confessed, or that this match is critical because it is the only strong evidence in the case. No information of this kind bears in any way on the actual forensic science inquiry, and risks creating an unconscious biasing effect on the examiner. Indeed, in most scientific fields, there is a careful and often formalized effort to shield researchers from this kind of contextual information.⁶¹ It's too dangerous. We human beings have a cognitive tendency to see what we expect to see. Think of the way that medical researchers make use, whenever possible, of carefully controlled studies to ensure that not even the treating physicians know who is receiving the medication under investigation and who may be getting the placebo. These

⁵⁹ OIG REPORT, *supra* note 52.

⁶⁰ See generally Stacey, *supra* note 20; Noelle Crombie, *FBI Blamed in Print Error*, THE OREGONIAN, Nov. 16, 2004, at A-1; OIG report, *supra* note 20.

⁶¹ D. Michael Risinger, et al., *The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1, 5 (2002).

protections exist to protect physicians from unconscious bias that might influence their interpretation of the effects of either the medicine or the placebo. Currently, in the forensic sciences, there are generally no such procedures to protect examiners from extraneous information that may have an unconscious influence on their findings.⁶²

To be sure, some information, though potentially biasing, may nonetheless be necessary for conducting the forensic test. A fingerprint examiner, for example, will likely need to know what surface a print came from, notwithstanding that the information may provide context clues about the crime itself. The point is not that examiners should lack all access to non-forensic information relating to the case. Rather, to the maximum extent practicable, they should *only* be given the case-related information that is actually relevant and helpful to their forensic inquiry. Dan Krane and others have coined the name “sequential unmasking” as a label for this approach, in order to emphasize that forensic analysts ought to learn only that information that they actually need, and only when they actually need it. All information, in other words, should be unmasked—that is to say revealed—to the examiner in sequence, and only when it is necessary.⁶³ The examiner should have access to all the information necessary to do his or her analysis effectively—no more, and no less.

The concern about the danger and power of biasing information is not simply theoretical. In a clever experiment, cognitive psychologist Itiel Dror used the Mayfield case to show the possibility of contextual bias effects on fingerprint examiners’ interpretations. A small handful of fingerprint examiners were each given a pair of prints, a latent print and a potential source print, and told that they were the prints from the Mayfield case. Each examiner was asked to evaluate whether or not the prints matched, using only the information contained in the print.⁶⁴

⁶² See generally *id.*; see also, e.g., Michael J. Saks et al., *Context Effects in Forensic Science*, 43 SCI. & JUST. 77 (2003).

⁶³ Dan Krane et al., *Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation*, 53 J. FORENSIC SCI., letter to the editor, 1006 (2008); see also Risinger et al., *supra* note 62; William C. Thompson, *Subjective Interpretation, Laboratory Error and the Value of Forensic DNA Evidence: Three Case Studies*, 96 GENETICA 153, 167 (1995).

⁶⁴ Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT’L, 74, 76 (2006).

In fact, however, unbeknownst to the examiners, the prints were *not* the Mayfield prints. Each examiner was actually given a set of prints that he or she personally had previously testified in court were a 100% certain, positive, error-free individualization. But now, when provided with this biasing contextual information suggesting that the prints were those involved in the Mayfield scandal, 60% of the examiners (three of the five examiners tested) reached *the opposite conclusion*, determining that the two prints in front of them did not in fact match.⁶⁵ A fourth examiner judged the prints to be inconclusive.⁶⁶ Only one of the five examiners reached a conclusion consistent with his or her original judgment that the prints matched.⁶⁷ To be sure, the Mayfield incident was a significant scandal, so the potential biasing effect of this context information was obviously quite extreme. Nonetheless, given some fingerprint experts' insistence that their methodology is not vulnerable to unconscious bias or general human fallibilities, Dror's findings generated a great deal of interest and a certain amount of both surprise and anxiety within the fingerprint community.⁶⁸ The experiment was, in a sense, a possibility proof, showing that bias could indeed, at least in some circumstances, be significant enough to affect examiners' conclusions. Follow up experiments by Dror and his collaborators on a larger number of examiners and with less starkly biasing information still revealed the potentially biasing effect of contextual information on analysts' judgments.⁶⁹

From one perspective, these findings are quite unsurprising. Research across a variety of other fields shows that we are *all* potentially biased by context and expectation.⁷⁰ Given that, why should we be in the least surprised that forensic science is no different than other cognitive enterprises?

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ *Id.*

⁶⁸ Sue Russell, *Bias and the Big Fingerprint Dust-Up*, MILLER-MCCUNE, June 18, 2009, available at <http://www.miller-mccune.com/legal-affairs/bias-and-the-big-fingerprint-dust-up-3629/>.

⁶⁹ Itiel E. Dror & David Charlton, *Why Experts Make Errors*, 56 J. FORENSIC IDENTIFICATION 600, 606, 612-14 (2006); Itiel E. Dror & Robert Rosenthal, *Meta-Analytically Quantifying the Reliability and Biasability of Forensic Experts*, 53 J. FORENSIC SCI. 900 (2008); see also Schiffer & Christophe Champod, *The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingerprint Individualization*, 167 FORENSIC SCI. INT'L, 116 (2007).

⁷⁰ See generally Risinger et al., *supra* note 62, and the many sources cited therein.

The recognition that fingerprint examiners are potentially subject to bias does not mean that they are behaving unprofessionally, or being careless, or not trying hard enough. Cognitive biases are an inherent danger of our cognitive architecture.⁷¹ That forensic scientists are not immune to them is hardly a surprise—except, perhaps, to those forensic scientists who were committed to a conception of their infallibility.

Dror's studies do make absolutely clear that bias is not simply a theoretical concern but a practical one as well. Moreover, they reveal that at least in some circumstances, biases may be acute enough to affect forensic examiners' outcome judgments. These studies therefore suggest both the theoretical and practical importance of taking concrete steps to limit examiners' access to biasing information. To be sure, these studies are preliminary, and involve only a relative handful of examiners. Nonetheless, these studies, coupled with all that is already known about bias in other cognitive domains, strongly suggest that this area should receive significant further study, so that we can better understand how and when bias poses a danger, and how often bias may be strong enough to affect an examiners' conclusions. There is also a significant need to identify the mechanisms that could usefully reduce those biasing effects that cannot be eliminated.

Furthermore, as law professor Michael Risinger has pointed out, the limited forensic science research on biasing effects to date ought not to be taken as an excuse for inaction.⁷² We know enough right now that we ought not to require further research before taking all reasonable action to reduce bias effects. The existing forensic-oriented studies, albeit limited in number and preliminary, coupled with the far more substantial research in other domains revealing human beings' cognitive vulnerability to bias, should reverse the burden of proof: unless the forensic science community can establish that it does *not* need blinding protocols, masking procedures, or other mechanisms that would reduce or eliminate bias, we should assume that bias-reduction mechanisms are indeed already warranted by our current degree of knowledge.⁷³

⁷¹ *Id.*

⁷² D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Glass Nine-Tenths Full (This Is About the Other Tenth)*, 50 JURIMETRICS J. 21 (2009) [hereinafter Risinger, *A Glass Nine-Tenths Full*].

⁷³ *See generally id.*

II. THE NATIONAL ACADEMY REPORT ON THE STATE OF FORENSIC SCIENCE

For those who closely follow the debates in forensic science, much of what I have described up to this point will have been quite familiar. But for those who began reading this Article believing in the television version, so to speak, the significantly more complicated reality may have come as something of a surprise. I hope I have, in this whirlwind tour, left those of you who walked with great confidence in the reliability of fingerprints at least a bit unsettled. Unfortunately, for the most part, the same cannot be said for the courts.

Beginning around a decade ago, enterprising attorneys began bringing admissibility challenges to fingerprinting, drawing to the courts' attention the kinds of problems and weaknesses I have just described. But nearly all of these challenges failed. One lone court did actually exclude fingerprint evidence in 2007, calling it "a subjective, untested, unverifiable identification procedure that purports to be infallible."⁷⁴ But although that case parroted what had practically become conventional wisdom in certain academic circles, as a legal decision it was a voice in the wilderness. Every other trial judge who has considered the admissibility of latent fingerprint identification since *Daubert* has found it to meet the applicable standard for the admissibility of expert evidence. It has been, for the most part, too hard for judges to contemplate excluding a form of evidence that has been routinely used for nearly a hundred years. For those academics engaged in constructive criticism of the forensic sciences, reading these judicial opinions has often felt like walking in a house of mirrors. As I will describe in more detail below, evidence of the problems facing forensic science would often be ignored, distorted, or recast, in order to help courts avoid

⁷⁴ Maryland v. Rose, No. K06-0545 (Balt. County Cir. Ct. 2007), available at <http://www.latent-prints.com/images/State%20Of%20Maryland%20vs%20Bryan%20Rose%20K06-0545.pdf>. However, that decision did not stand. A federal judge overturned the admissibility call, saying "fingerprint identification evidence based on the ACE-V methodology is generally accepted in the relevant scientific community, has a very low incidence of erroneous misidentifications, and is sufficiently reliable to be admissible under FED. R. EVID. 702 generally and specifically in this case." United States v. Rose, No. CCB-08-0149, 2009 WL 4691612, at *3 (D. Md. Dec. 8, 2009).

confronting the insufficient research basis supporting these forms of evidence.⁷⁵

Meanwhile, late in 2005, partly at the request of the leadership of the forensic science community itself, Congress commissioned the well-regarded, independent, and non-partisan National Academy of Sciences to research and write a report on the needs of the forensic science community. In February of 2009, the long-anticipated report was issued.⁷⁶

This report was written by an interdisciplinary panel of distinguished scholars and practitioners, who conducted their own investigation into the state of the research, and also heard numerous days of testimony from a substantial number of leading forensic science professionals, researchers, and others knowledgeable about the state of the forensic scientists.⁷⁷ This panel included scientists from a variety of fields, several forensic professionals, and some with legal experience.⁷⁸ The panel also included one law professor: Brooklyn Law School's own Margaret Berger, in whose honor this Festschrift volume of the Law Review has been written. In essence, the 319-page report substantially confirms the views of the academic critics about the inadequacy of the research basis to support many of the claims routinely made by forensic scientists.⁷⁹

For example, the report finds that there is not an adequate basis for claims of individualization. The report also finds "a notable dearth of peer-reviewed, published studies establishing the scientific bases and validity of many forensic methods."⁸⁰ The report claims that research on proficiency, performance, and the role of bias and observer effects is "sorely needed."⁸¹ All in all, "[t]he present situation . . . is seriously wanting, both because of the limitations of the judicial system

⁷⁵ See *infra* Part III.

⁷⁶ NAS Report, *supra* note 9.

⁷⁷ In the interests of full disclosure, I should note that I provided oral commentary to the NAS panel during their process.

⁷⁸ The make-up of the committee has been criticized by the forensic science community for having insufficient practitioners. It is interesting to note that the committee in fact had several forensic practitioners, while it had not a single member who had already published critical work concerning the adequacy of the research basis of forensic science. For more information on the committee and its make-up, see generally D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Path Forward Fraught with Pitfalls*, UTAH L. REV. (forthcoming).

⁷⁹ NAS Report, *supra* note 9.

⁸⁰ *Id.* at 5-6.

⁸¹ *Id.*

and because of the many problems faced by the forensic science community.”⁸²

The report’s boldest and perhaps most important recommendation is for the creation of a new independent federal agency to regulate, supervise, and improve the forensic sciences. This agency, dubbed the National Institute of Forensic Science (NIFS), would be responsible for funding research to improve forensic sciences; it would also be responsible for establishing and developing best practices, and, more generally, supporting and overseeing the forensic science infrastructure.⁸³ One academic has quipped that the NIFS would be “a mixture of the SEC and the NSF” for forensic science, rather an apt characterization.⁸⁴ The report explains in detail that no other existing agency has, in the committee’s view, the ability effectively to provide all of what forensic science needs in terms of both research management and regulatory oversight.⁸⁵ No existing agency—neither NIST, the NIJ, nor anyone else—has, according to the report, the necessary expertise, resources, and appropriate political culture to permit it to perform this array of functions credibly and successfully.⁸⁶

The report makes a number of other significant recommendations. It calls in strong terms for additional research to establish the validity and reliability of forensic sciences, as well as research to examine the extent of biases and observer effects.⁸⁷ It calls for mandatory laboratory accreditation and mandatory individual certification of forensic scientists (right now both are entirely optional).⁸⁸ Significantly, it calls for the use of incentive funding to motivate states to make their crime laboratories independent from law enforcement and prosecutors.⁸⁹

The report has received a good deal of attention both from within the forensic science community and from outsiders. As of now, several hearings on Capitol Hill have been held to

⁸² *Id.* at 5-9.

⁸³ *Id.* at 5-14.

⁸⁴ Roger Koppl, Professor of Econ. & Fin., Farleigh Dickinson Univ., Remark at a Forensic Science Conference at Arizona State University Sandra Day O’Connor College of Law (2009).

⁸⁵ NAS Report, *supra* note 9, at 5-24.

⁸⁶ *Id.*

⁸⁷ *Id.* at 5-18.

⁸⁸ *Id.* at 5-19.

⁸⁹ *Id.* at 5-17.

consider its recommendations.⁹⁰ President Obama has created a forensic science task force to consider how to proceed.⁹¹ But the unfortunate reality is that now, roughly a year after the report was issued, almost no one believes that NIFS is going to be on the horizon any time soon. A new federal agency seems to be neither fiscally nor politically viable.

Overall, the NAS report is an impressive achievement, and both its criticisms and suggestions are, in my view, generally on the mark. However, I do have one relatively significant quarrel with the report, which relates to its treatment of the legal aspects of forensic science and admissibility. The report offers a thorough and trenchant critique of how the courts have thus far handled forensic science. It describes how the courts have substituted long use for an actual focus on proven validity.⁹² It points out the ways in which judges have been “utterly ineffective” at honestly assessing the research basis of the pattern identification sciences.⁹³ I agree completely. The report further contends that the judiciary, particularly given judges’ lack of training in science, the case-by-case nature of the adjudicatory system, and the limits of appellate review, cannot be expected to solve this problem on its own. “Judicial review, by itself, will not cure the infirmities of the forensic science community.”⁹⁴ Again, I wholeheartedly agree.

But then, at this key point, the NAS report decides to punt. After offering this significant critique of the judiciary’s actions, the report is distressingly silent about what the judiciary ought to do next. If judicial review by itself will not and cannot solve our problems in the forensic science arena, does it nonetheless have some role to play? While we await NIFS, or any other significant federal initiative vis-à-vis forensic science; while we await the necessary research that we hope will eventually be done; while we await greater regulation and the possibility of mandatory accreditation, what should

⁹⁰ *Strengthening Forensic Science in the United States: Hearing Before the S. Comm. on the Judiciary*, 111th Cong. (2009), webcast available at <http://judiciary.senate.gov/hearings/hearing.cfm?id=4038>. The Congressional Subcommittee on Technology and Innovation also held a hearing on March 10, 2009 and the Judiciary Committee also held a hearing on May 13, 2009.

⁹¹ The National Science and Technology Council in the Office of the President of the United States created a Subcommittee on Forensic Science. For details, see <http://projects.nfstc.org/trace/2009/presentations/7-melson-stolorow-nas.pdf>.

⁹² NAS Report, *supra* note 9, at 5-9.

⁹³ *Id.*

⁹⁴ *Id.* at 12.

judges do? Sit idly by, and continue permitting pattern identification evidence as they always have? While providing a good deal of persuasive authority to any judge who wishes to scrutinize forensic science with more care, the NAS report does not offer any specific guidance whatsoever for courts wrestling with admissibility determinations under *Daubert*.

Understandably, the committee may have thought that sweeping statements about admissibility were unwarranted, and more generally that telling judges how to behave would be overstepping its mandate. I fully recognize that the NAS report was directed at Congress and was not primarily intended for the courts. But the committee might have done more. For example, while the report says there is no scientific basis for claims of individualization, it does not go one step further and explicitly say that such testimony ought therefore to be deemed objectionable.⁹⁵ The report says that a zero error rate is not scientifically plausible, but it does not explicitly say that testimony asserting such a rate should therefore be viewed with skepticism.

To be sure, these are easy inferences to draw from what the report *does* say. If *Daubert* says that expert evidence needs to be established as reliable and valid, and a court accepts the NAS report's conclusion that a zero error rate is not scientifically plausible, then it might be a simple matter of syllogistic logic to conclude that any claim of a zero error rate does not pass *Daubert*. And there is no doubt that the report is already being cited by defense attorneys in the latest round of *Daubert* motions that are, inevitably (and, in my view, quite appropriately) being spurred by the report's contents and conclusions.

But there is, in the report, a certain sense of resignation, perhaps even fatalism, about the courts. There is a tone that suggests that the committee may have thought that the judiciary has done such a poor job of gatekeeping in this area that it is hard to imagine the possibility of change.⁹⁶ To have faith in the courts' capacity to make an about-face in this domain is, perhaps, the mark of a naïve optimist, or maybe even a chump.

While I agree with the report that courts alone cannot and will not prompt the necessary reforms to forensic science,

⁹⁵ *Id.*

⁹⁶ *See generally id.*

it also seems to me that courts absolutely do need to be *part* of the solution. For courts to continue to treat forensic science evidence in the same manner that many of them did prior to the NAS report would be worse than cowardly. The NAS report strongly suggests that the concerns expressed in recent years by a number of academic critics were neither overstated nor illegitimate. The academic critics were not crackpots, nor were they Chicken Littles worrying about something that no one else could see because it wasn't really there. There truly is an insufficient basis in research for many of the strong claims that forensic scientists have been making for years. The courts should squarely confront that fact and decide in a thoughtful way what consequences it creates for these forms of evidence; what effects it has on admissibility; and whether it means that the evidence should, at a minimum, be limited, and if so, how.

Admittedly, the question of whether fingerprint evidence should pass *Daubert* or the equivalent test under state law is a difficult one. Fingerprint evidence has been used in court for roughly one hundred years. There is no doubt that the pattern variation among human friction ridges is indeed enormous, and that fingerprints, whether or not they are truly unique, certainly have extremely significant discriminatory power. Fingerprint evidence is quite obviously probative. Moreover, people trust it. Not just “people”—not just the proverbial man in the street, not just forensic scientists themselves, not just prosecutors. Even those of us—like me—who are sometimes viewed as critics of fingerprint evidence acknowledge its probative power.⁹⁷ For example, if you asked me which piece of evidence I would have more confidence in, an eyewitness identification by a crime victim of the perpetrator, a stranger who was viewed for a short period of time by the victim during a stressful crime; or a latent fingerprint identification of an individual made from several high-quality, clear latent prints found in a location and in circumstances strongly suggesting that whoever left the print was connected to the crime, my honest answer would be the fingerprint identification. All of these facts make it extremely difficult for a court to seriously and deeply consider the possibility of excluding, or even limiting, this form of evidence.

⁹⁷ Indeed, even Simon Cole and Michael Saks, two of the staunchest so-called critics, both recognize that fingerprint evidence has a great deal of power. See, e.g., Cole, *Grandfathering Evidence*, *supra* note 1, at 1193-94; Saks, *Merlin and Solomon*, *supra* note 1, at 1106.

And yet, in a way these facts about the power of fingerprint evidence are beside the point, or at least they do not adequately and fully answer the question of fingerprint evidence's legitimate admissibility in court. The problem with fingerprint evidence is *not* that it completely lacks probative power, but rather that research on the domain has not yet established the appropriate *limits* to its probative power, or shown how that value varies depending on its quality or its quantity of information. It is as if I wanted to sell you a valuable jewel and the question was how much it should cost. We may both agree that the jewel is indeed valuable, but the right question is not whether it has value, but just *how* valuable it is, and just how we can know its *appropriate* value. If I refused to research that question, or even to permit any research on the specific characteristics of the jewel that would help assess its precise value, and I further refused to research the relevant market information to help determine its appropriate price, I would have an extremely hard time selling the jewel, and legitimately so. If all I said was, "Of course it's valuable. We both know it's simply the most valuable jewel in the world. I'm not going to investigate the question of its precise value, but we both know it's virtually perfect. Let's therefore just agree that it's worth an extraordinary amount and price it accordingly," we would both expect you to walk away from the potential sale.

Fingerprint evidence, like that jewel, is obviously valuable. But like that jewel, we should be wary of 'buying' it as legal evidence in court until we have a better, research-based understanding of precisely how valuable it is, and whether there are instances in which we might be assuming it to have a significantly higher value than it really does. Without careful proficiency testing of examiners, without information about what the significance of any given 'match' really is, without error rate information about the frequency and circumstances of mistakes, without understanding which fingerprint identifications are easy and which ones are more difficult and hence more likely to be error-prone, we should be cautious buyers indeed. Quite possibly, the better strategy might be simply to forego the purchase until more of the necessary information is available. Still, recognizing this somewhat awkward state of affairs—that we have a form of evidence that obviously often has probative power, but also has a strikingly inadequate research basis—helps, I think, to make sense of the courts' reluctance to exclude, or in some

circumstances, even honestly to engage with the arguments made by defense counsel.

III. HOW HAVE THE COURTS ANALYZED THE ADMISSIBILITY OF FINGERPRINTS?

In the remaining portion of this article, I will explore both what courts have actually done when confronted with challenges to the admissibility of pattern identification evidence, as well as what they ought to do. First, I will focus on what they have done, and then turn to what they ought to do, but it may be valuable to preview my suggestions about what they ought to do before turning to look at how courts have actually responded to the numerous pattern identification admissibility challenges that have arisen.

My bottom line is straightforward: *Forensic science experts should not continue to be given free rein to testify in the manner they have typically done up until now. Judges need to develop a variety of thoughtful approaches—a toolkit of sorts—with which they can assess admissibility, and this toolkit should absolutely include outright exclusion in some circumstances.* What judges ought to do may well not be the same across the board, even within the same field—as *Kumho Tire v. Carmichael*⁹⁸ indicates, and as Michael Risinger has usefully emphasized,⁹⁹ the court’s responsibility when assessing the admissibility of expert evidence is to focus on “the task at hand”—which means looking closely at the *specific* nature of the claims being made in the *particular* circumstance. This means that we may not be able to achieve—nor may we necessarily want—a one-size-fits-all admissibility approach to any given form of pattern identification evidence. Admissibility determinations, and the scope of permissible testimony, may depend on the details—is it a single partial latent from AFIS, or are there nine extremely clear latents corresponding to three different fingers of the defendant? What is the quality of the print or prints in question? What specific claims is the expert trying to make, and how absolute are the conclusions presented? These questions, and others like them, ought to inform judges’ analyses. All pattern identifications are not

⁹⁸ 526 U.S. 137 (1999).

⁹⁹ See generally D. Michael Risinger, *Defining the “Task at Hand”: Non-Science Forensic Science After Kumho Tire Co. v. Carmichael*, 57 WASH. & LEE L. REV. 767 (2000) [hereinafter Risinger, *Defining the “Task at Hand”*].

created equal, and a blanket approach to admissibility, pro or con, is not likely to be warranted.

Nonetheless, at a bare minimum, given currently available knowledge, courts ought not to permit evidence of either individualization or a zero error rate. Even the forensic science community is beginning—slowly and in the case of individualization somewhat falteringly—to recognize that these claims are perhaps better let go.¹⁰⁰ But I would go further. In many cases involving pattern identification, courts have, as of now, only two legitimate choices if they are to take *Daubert* seriously: either (1) limiting the evidence by restricting it to description of similarities and differences, rather than offering opinions; or (2) outright exclusion. One alternative is to greatly restrict the expert's testimony, limiting it to description of similarities and differences without any evidence providing an interpretation of these similarities and differences. This option, though superficially appealing, raises some thorny admissibility concerns which are hard to avoid in a principled way, as I will describe below. Nonetheless, it seems acceptable as an imperfect, interim solution for courts who simply find it too hard to exclude evidence that likely does have significant probative value. The second approach, and one that deserves far more serious consideration than most courts have been prepared to give it, is outright exclusion.

While that may sound like an extreme reaction—cutting off the patient's arm when perhaps it could have been put in a cast and saved—it is important to note, as I will argue below, that with a modicum of effort on the part of researchers and forthright cooperation from the forensic science community, exclusion would be quite short-lived. As I will argue in Part IV,

¹⁰⁰ On February 19, the President of the IAI, Bob Garrett, issued a two-page statement to its membership stating, "Although the IAI does not, at this time, endorse the use of probabilistic models when stating conclusions of identification, members are advised to avoid stating their conclusions in absolute terms when dealing with population issues." Though the wording is perhaps slightly opaque, this is, in essence, a pulling back from an absolute claim of individualization. See http://www.theiai.org/current_affairs/nas_memo_20090219.pdf (last visited April 17, 2010). However, a month later, in written comments prepared for the Senate Judiciary Committee, the IAI backed off from this claim, and in a significantly longer document (7 pages of remarks), pointedly said nothing about individualization, whether conclusions can be absolute, or how to think about population issues. See http://www.theiai.org/current_affairs/nas_response_leahy_20090318.pdf (last visited April 17, 2010). For another example of a staunch defender of the reliability of pattern identification science suggesting that, for instrumental rather than epistemic purposes, claims of individualization should be dropped, see John Collins, *Stochastics—The Real Science Behind Forensic Identifications* (2009), Crime Lab Report, available at http://www.crimelabreport.com/library/monthly_report/11-2009.htm.

courts' primary focus under *Daubert* should be on the question of whether the expert's claims have been subject to adequate *testing*—whether there is evidence that supports the claim that the expert can actually do what she says she can do. Complete knowledge of the cognitive practices of fingerprint evidence; a validated statistical model of ridge characteristics and frequencies; even objective standards for determining whether or not prints match—all of these forms of knowledge would be valuable, beneficial, and are absolutely worth pursuing, but they need not exist in order for latent fingerprint evidence to be legitimately admissible under *Daubert*. What courts really ought to consider requiring, in many cases, as a *minimum prerequisite to admissibility* is simply much better error rate information about examiners' abilities in practice. And producing this information is eminently achievable with concerted focus and effort.

I turn, now, to what courts have actually done when confronted with admissibility challenges to fingerprint evidence and other kinds of pattern identification evidence. Up to the present, courts wrestling with these admissibility challenges have offered several categories of arguments and made certain repeated intellectual moves. In this Part, I will describe three dominant analytic approaches that courts have taken: (1) the ostrich maneuver, (2) the ACE-V conclusion (where ACE-V stands, in this case, for “Admissible—Considering Everything, it's Valid (enough);” and (3) the Solomonic compromise.

A. *The Ostrich Maneuver: Problem? What Problem?*

The first approach taken by some courts when confronting challenges to the admissibility of pattern identification evidence is what I call the ostrich maneuver, because these courts appear to be trying desperately to keep their heads in the sand. Certain judges have, to a sometimes remarkable extent, averted their eyes to the quite legitimate concerns about the research basis supporting the conclusions offered in the pattern identification sciences. These judges have more or less asked, “Problem? What problem?” An archetypal example of the ostrich maneuver occurred in *United States v. Havvard*.¹⁰¹ The judge's view of the issue was captured early in

¹⁰¹ 117 F. Supp. 2d 848 (S.D. Ind. 2000). Though *Havvard* is an especially dramatic “ostrich” opinion, it is certainly not the only one. See, e.g., *United States v.*

his written opinion finding that fingerprinting passed *Daubert*, when he wrote, “The court’s decision may strike some as comparable to a breathless announcement that the sky is blue and the sun rose in the east yesterday.”¹⁰²

The judge went on to explain why, in his opinion, fingerprint evidence posed virtually no genuine difficulty under *Daubert*. He concluded, for example, that fingerprint evidence could be relied upon because it had indeed been tested (testing being one of the most important factors under *Daubert*) “in adversarial proceedings with the highest possible stakes—liberty and sometimes life.”¹⁰³ Note the judge’s slippage here—*Daubert* envisions *scientific* testing, not *courtroom* testing. In fact, the entire purpose of *Daubert*—a heightened reliability screen for expert evidence—derives from the idea that the crucible of the courtroom is, by itself, an insufficient check on validity and reliability for scientific and expert evidence. If testing through adversarial proceedings were enough, then a separate judicial inquiry into reliability would be entirely superfluous.¹⁰⁴ We wouldn’t need *Daubert* at all, because the crucible of the courtroom would suffice. Notwithstanding the

Reaux, 2001 WL 883221 (E.D. La. 2001); *State v. Cole*, 2002 WL 1397452 (Del. Super. 2002); *United States v. Rogers*, 26 Fed. Appx. 171, 2001 WL 1635494 (C.A.4 (N.C.) 2001).

¹⁰² *Havvard*, 117 F. Supp. 2d 848 (S.D. Ind. 2000).

¹⁰³ *Id.* at 854.

¹⁰⁴ To be sure, one could make the argument that we do not actually need a special judicial check on validity and reliability for expert evidence. One could possibly defend this claim on a variety of grounds, ranging from (1) an argument that adversarial testing through the presentation of contrary evidence and cross-examination are in fact an adequate method for evaluating expert evidence; to (2) that there may be little evidence that judges’ evaluation of the legitimacy of expert evidence is epistemically superior to that of juries, in which case using them as a ‘reliability’ screen for evidence might be epistemically ineffective and, therefore, potentially usurping the power of the jury for no legitimate purpose; to (3) that we do not engage in strong validity checks on other kinds of potentially unreliable evidence and there is an insufficient institutional or epistemic justification for treating expert knowledge differently, etc. But none of these arguments—whether or not they have any intellectual merit—provide a legitimate basis for a lower court within our legal hierarchy to argue that ‘adversarial’ testing constitutes testing of the sort envisioned by *Daubert*. A judge could make a more subtle argument that adversarial testing is adequate for “shaky but admissible” evidence. *Daubert* says, and indeed the judge in *Havvard* quoted, later in his opinion: “Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.” *Havvard*, 117 F. Supp. 2d at 854 (internal quotation marks omitted) (quoting *Daubert*, 509 U.S. at 596). But while the power of adversarial testing might be an argument for permitting borderline evidence (and indeed, is a contributing argument to the ACE-V approach to these questions discussed below, *infra* notes 114-128 and accompanying text), adversarial testing should not be understood as the kind of testing referred to and expected by *Daubert*.

sleight of hand at work in *Havvard* with respect to the idea of testing, adversarial testing simply cannot legitimately be seen as fulfilling *Daubert's* idea of testing.

The judge engaged in a similar, though less egregious, sleight of hand with respect to the peer review and publication factors of *Daubert*. He granted that the “publication” factor was an awkward fit for fingerprint evidence because it had been developed “for forensic purposes,” but found that adversarial testing offered an adequate substitute.¹⁰⁵ He also found that there was plenty of peer review, both because of the practice of having one examiner verify the conclusions of another, and also because any examiner could review the conclusions of any peer by taking her own look at the prints in question.¹⁰⁶

Is the verification stage of ACE-V akin to peer review? It is true enough that having verification as a standard practice does provide a certain degree of routine peer examination for each declared fingerprint match. But peer review of, for example, manuscripts for publication, usually, though not always, makes the reviewers blind to authorship. More often than not, peer reviewers do not know whose manuscripts they are reading so that they are not potentially biased by the authors' credentials or experience. Moreover, peer review in the context of research assessments permits a kind of semi-public scrutiny and dialogue regarding the merits of the researchers' methodology and approaches, precisely because it typically involves scrutiny by other experts from *outside* the original author's workplace or close circle of collaborators. Having your immediate colleagues take a look at an article is not what journals—or *Daubert*—means by peer review. Moreover, peer review usually also requires written commentary and reasoned explanation.¹⁰⁷

The verification phase of fingerprint evidence thus fails to map precisely onto scientific peer review in several important ways. To be sure, if verification were conducted as a “blind” review, i.e., the verifying examiner did not know whose identification she was verifying or even that she was conducting a verification rather than an initial analysis, it

¹⁰⁵ *Havvard*, 117 F. Supp. 2d at 854.

¹⁰⁶ *Id.*

¹⁰⁷ On peer review, see generally Effie J. Chan, *The 'Brave New World' of Daubert: True Peer Review, Editorial Peer Review, and Scientific Validity*, 70 NYU L. REV. 100 (1995); Lars Noah, *Scientific "Republicanism": Expert Peer Review and The Quest for Regulatory Deliberation*, 49 EMORY L.J. 1033 (2000); Susan Haack, *Peer Review and Publication: Lessons for Lawyers*, 36 STETSON L. REV. 819 (2007).

would be a closer, though still not exact, fit. But even blind peer review would be more of a check on consistency—reliability in the scientific, rather than legal sense—rather than validity or accuracy. That two examiners reached the same conclusion shows no more and no less than that they reached the same conclusion. If blind verifiers consistently reach the same conclusion as the original analyst, it shows that the method generates *consistent* results, but it does not necessarily show that it achieves *accurate* results, though in some instances we may be willing to infer accuracy from consistency.¹⁰⁸

Clearly, then, verification cannot simply be equated with academic peer review. However, I grant that equating verification with peer review is not quite as distressing, conceptually, as equating adversarial testing with scientific testing, especially given that *Daubert* intended its criteria as suggestions rather than hard requirements. These suggestions may be appropriately modified to apply to the particular circumstances of the practices at issue and the “task at hand” in the specific case¹⁰⁹ and thus a loose analogy between peer

¹⁰⁸ Accuracy and consistency are not necessarily correlated—I could have a miscalibrated instrument that gives me the same result every time and yet is always wrong. It is equally possible that there could be a method that generates consistent results even when practiced by different individuals, but these consistent results were not accurate. Think for example, of a poorly designed psychological test for depression. Different testgivers might use the test on patients, get consistent diagnoses, and yet be wrong, because the test wasn’t well designed in the first place. Ironically, in the case of latent fingerprint identification, the absence of a consistent method (given that, as discussed above, at *infra* notes 22-23 and accompanying text, ACE-V is less a method than an outline of procedural steps for careful looking) might legitimate an inference of some degree of validity from generally consistent results through blind peer review. If examiners have extremely consistent results notwithstanding the fact that they use an array of imprecisely defined and not-quite-identical approaches, including choosing different minutiae, different standards for determining the quantum of information required to call an identification, and varied practices for conducting a comparison, we might be prepared to think their consistent results provided some warrant for inferring validity. It is as if instead of conducting the same psychological test multiple times, we conducted several different psychological tests and all of them resulted in the same diagnosis; these multiple methods’ consistent results would be cumulative, and thus could increase our confidence in the correctness of the original diagnosis. The irony is that this potential evidentiary power of verification results from the lack of formalized or consistent standards; to whatever extent examiners are doing the same thing as one another, verification can establish only consistency, or a check on methodological error, but not validity, because the verification evidence would be merely redundant, rather than cumulative. On cumulative and redundant evidence, see generally DAVID A. SCHUM, *THE EVIDENTIAL FOUNDATIONS OF PROBABILISTIC REASONING* (2001). For an analogous argument in the context of human/technology partnerships in fingerprint evidence, see Dror & Mnookin, *supra* note 17, at 60-65.

¹⁰⁹ *Kumho Tire*, 526 U.S. 137, the third case in the so-called *Daubert* trilogy, emphasizes the importance of focusing on the “task at hand.” For discussion of the

review and verification might not be out of place. But given that there is absolutely no publicly-available data about how often peer review of the sort engaged in by fingerprint examiners actually catches errors, or how well it functions in operation, the courts' easy willingness to presume its utility seems to be overreaching. For example, if non-blind verifiers virtually never disagree with the original analyst, then the power of this "review" as a check may in fact be quite limited, for it might be that whatever errors do occur are, as in the Mayfield case, likely to be repeated by subsequent analysts as well.¹¹⁰

But while the analysis in *Havvard* is generally unpersuasive, it is the opinion's conclusion that is most ostrich-like of all. The court wrote, "In sum, despite the absence of a single quantifiable standard for measuring the sufficiency of any latent print for purposes of identification, the court is satisfied that latent print identification easily satisfies the standards of reliability in *Daubert* and *Kumho Tire*. In fact, after going through this analysis, the court believes that latent print identification is *the very archetype of reliable expert testimony under those standards*."¹¹¹ Reasonable people can disagree about whether fingerprint evidence ought to be admissible under *Daubert*. But to argue that it is the "very archetype of reliable expert testimony under those standards"¹¹² strains all credibility. Only by putting one's head in the sand could one possibly conclude that latent fingerprint evidence—which has been tested in the adversarial crucible but not scientifically, lacks meaningful error rate information, and operates without statistical foundation or any validated, objective criteria for determining a match—is the archetype of reliable evidence under *Daubert*.¹¹³

importance of this focus, see generally Risinger, *Defining the "Task at Hand"*, *supra* note 99.

¹¹⁰ However, if verifiers disagree with the original analyst more than occasionally, that provides indirect evidence that the error rate for an individual examination is not, in fact, negligible.

¹¹¹ *Havvard*, 117 F. Supp. 2d at 854 (emphasis added).

¹¹² *Id.*

¹¹³ Another ostrich maneuver, in some ways even more extreme than the superficial and poorly-reasoned analysis in *Havvard*, occurs when courts decide that pattern identification evidence is so clearly reliable that they do not even permit the defense to hold a preliminary hearing on the issue. See, e.g., *United States v. Joseph*, 2001 WL 515213, *1 (E.D. La. May 14, 2001), for one example. Of course, this issue is complicated by the question of what the defense actually proffers. For example, if a defense attorney asks for a hearing but has done no substantial preparation and fails to make a credible argument, a judge would appropriately deny the request. That

B. *The Judiciary's Own Version of ACE-V*

The second approach is what I, slightly tongue-in-cheek, call the judiciary's version of ACE-V. In this case, the acronym stands for "Admissible—Considering Everything, it's Valid (enough)." This category of cases includes those courts who attempt, with at least a modicum of seriousness, to engage *Daubert* and *Kumho Tire's* requirements vis-à-vis pattern identification testimony, and who recognize that the research basis supporting the evidence is not what one would wish it to be. Nonetheless, 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.

United States v. Sullivan is an illustrative example of this approach.¹¹⁴ The court wrote:

The court shares the defendant's skepticism that the ACE-V methodology enjoys a 0% error rate, making it effectively a perfect art. There is no evidence, however, that the ACE-V methodology as performed by the FBI suffers from any significant error rate. The FBI examiners have demonstrated impressive accuracy on certification-related examinations, and Younce testified that an examiner who made a false identification would be finished as an examiner due to the difficulty in rehabilitating him or her as a witness. While the defendant is correct that the party submitting the evidence has the burden of establishing its reliability under *Daubert*, the defendant has failed to submit any evidence to dispute the plaintiff's evidence of a minimal error rate. Consequently, while the court rejects the plaintiff's claim of a 0% error rate, it finds that the

appears to be what occurred, for example, in *United States v. Pena*, in which the trial court reasoned, "the case law is overwhelmingly in favor of admitting fingerprint experts under virtually any circumstance." 586 F.3d 105, 110 (1st Cir. 2009) (internal quotation marks omitted). "Consequently, the court reasoned, the only way it would have considered excluding the testimony or giving a limiting instruction 'is if there had been data, real evidence presented about the limitations of fingerprinting.'" *Id.* "Instead, as the court acknowledged, Pena's motion to exclude relied on 'one article from the Fordham Law Review, and that's not enough to carry the weight of the exclusion motion.'" *Id.* It is quite understandable that a court is not going to waste its time with a preliminary hearing when the defense is so ill-prepared. But the appellate case law in the case suggests that it is never an abuse of discretion to admit fingerprint evidence without a preliminary hearing, and this is a distressing—and ostrich-like—conclusion. *Id.* Interestingly, just as this Article was going to press, Judge Nancy Gertner, whose opinion refusing a *Daubert* hearing was affirmed in *Pena*, subsequently issued a novel procedural order explicitly referencing the NAS report, to make clear her openness to hearing *well-prepared* challenges to pattern identification evidence. See Procedural Order, available at <http://www.mad.uscourts.gov/boston/pdf/ProcOrderTraceEvidenceUPDATE.pdf> (last visited April 17, 2010).

¹¹⁴ 246 F. Supp. 2d 700 (E.D. Ky. 2003).

error rate is not sufficient to render ACE-V unreliable under *Daubert*.¹¹⁵

The court squarely recognized that a zero percent error rate is not credible. The court further acknowledged that the proffering party had the burden of showing reliability, which ought to have meant that a lack of error rate information counted against the prosecution.¹¹⁶ Nonetheless, in the face of this acknowledged absence of information supporting any particular error rate, the court found that good performance on certification exams and a culture that deems any discovered identification errors so impermissible as to be career-ending, spoke sufficiently to the question of error rate as to render the ACE-V approach to fingerprint identification reliable under *Daubert*. To be sure, in the right circumstances, proficiency tests could indeed provide substantial and adequate information regarding error rates, but the current certification tests are neither challenging enough nor taken in circumstances sufficiently mirroring actual practice for this inference to be warranted.¹¹⁷ Similarly, the fact that the professional culture within the latent fingerprint community may view any false identification error as career-jeopardizing if discovered does not itself establish that errors do not occur. Indeed, one could imagine that a professional culture that deems certain errors career-ending might also be a culture that chooses not to look very hard for them.

Turning to the testing factor of *Daubert*, the judge explained:

The court further finds that, while the ACE-V methodology appears to be amenable to testing, such testing has not yet been performed. The court disagrees that testing that establishes the validity of the principles underlying ACE-V—that fingerprints are unique and permanent—can substitute for testing of the ACE-V methodology itself. That testing, however, is relevant as it provides a foundation for the ACE-V methodology. . . . But as the defendant points out, there is not a standard defining how many similarities must be found before a match is declared. Younce testified that there is no minimum number of “points” in common necessary to declare a match between a known and an unknown print. Indeed, Younce testified that such a requirement would be unscientific. While it is possible that this position is ultimately correct, it is not supported by the studies submitted by the plaintiff. Evidence that no two

¹¹⁵ *Id.* at 704.

¹¹⁶ *Id.*

¹¹⁷ See generally Koehler, *Proficiency Tests*, *supra* note 1; Mnookin, *Confessions of a Fingerprinting Moderate*, *supra* note 1, at 136-37.

fingerprints are the same-or that no two 21.7% of a print are the same-is not evidence that no two fingerprints can share a partial print in common. The court finds that this concern does not render fingerprint evidence unreliable for the purposes of *Daubert*. While the possibility that two fingers may have a fractional portion of a print in common may affect the probability estimates that two fingers may leave the same fractional print, that possibility goes to the weight of the evidence, not its admissibility.¹¹⁸

Note that here, the court determined that testing *could* be done and that it *had not yet* been done. Furthermore, the court understood, quite correctly, that evidence supporting the uniqueness and permanence of fingerprints did not itself establish that fingerprint examiners could actually make certain identifications. The court further recognized the lack of objective standards for determining a match. But at this point, the court essentially punted. Without any serious effort at analysis or explanation, the court simply decided that these difficulties went to the weight of the evidence rather than its admissibility.¹¹⁹ Still, compared to ostrich judges, this court at least elected to name clearly some of the problems with the research basis supporting fingerprint identification.

Another one of the many examples of judicial ACE-V thinking is the appellate opinion in *United States v. Mitchell*.¹²⁰ The court acknowledged that if “directed, specific actual testing” were the requirement of *Daubert*, then fingerprint evidence would have significant problems.¹²¹ However, the court seemed to believe that “directed, specific, actual testing” was not in fact required; instead, the long and substantial history of the use of the technique could provide what the court considered to be a form of “implicit testing.”¹²² Moreover, the court emphasized that while it was indeed required to be a “gatekeeper” for scientific evidence, it was also “only a gatekeeper, and a gatekeeper alone does not protect the castle.”¹²³ All in all, though the evidence supporting validity might be imperfect, the real question, in the court’s estimation, was whether the available evidence of validity was sufficient to pass the baton from the court to the adversary system.¹²⁴ In other words, fingerprint evidence might not be Valid (with a

¹¹⁸ *Sullivan*, 246 F. Supp. 2d at 704.

¹¹⁹ *Id.*

¹²⁰ 365 F.3d 215 (3d Cir. 2004).

¹²¹ *Id.*

¹²² *Id.*

¹²³ *Id.*

¹²⁴ *Id.*

capital V) but it was valid enough, in the court's estimation, for cross-examination and counter-expertise to be an adequate corrective to whatever limitations might be present.¹²⁵

Another, and perhaps the best-known, example of judicial ACE-V thinking is Judge Louis Pollak's second opinion in *United States v. Llera Plaza* ("*Llera Plaza II*").¹²⁶ The first time he considered fingerprint evidence, the judge—an especially thoughtful jurist, and a former professor and dean at Yale Law School—made fingerprint history by ruling, in *Llera Plaza I*, that latent fingerprint evidence did not pass muster under *Daubert*. After a re-hearing on somewhat unusual procedural grounds, Judge Pollak, by his own admission, changed his mind. In his second opinion, *Llera Plaza II*, he continued to find that *Daubert's* testing factor was not fulfilled. On error rates, he found that the extant proficiency tests gave little assistance because they were too easy to provide a

¹²⁵ There is an important point lurking here. The best argument for admissibility under *Daubert* is that the problems with fingerprint evidence are problems that can adequately be made clear and explicated to a lay factfinder on cross-examination. Hence the evidence need not be excluded, because, given that this isn't "rocket science," as Judge Gertner put it in an important case limiting handwriting identification evidence, juries can adequately understand the weaknesses and evaluate accordingly. *United States v. Hines*, 55 F. Supp. 2d 62, 70 (1999). Whether or not this is a fair assessment of jury capacity, it is far from clear why the adversary system's mechanisms should be deemed adequate for forensic science, but not in other arenas of expert evidence. In any event, for this argument to have any merit whatsoever, judges do clearly have to give the defense significant leeway in calling expert witnesses who can provide these important field critiques. There have been several judges who, having decided that the evidence is admissible under the applicable legal standard, have then refused to allow the defendant to offer testimony challenging its fundamental validity. This is both disturbing and disingenuous—if the best argument for admission of pattern identification evidence is that it can be adequately "impeached" through the adversary process, then it is imperative that the defense have the opportunity to present the weaknesses and uncertainties of the evidence. For examples of cases that undertake this sleight of hand, see, e.g., *State v. Armstrong*, 920 So. 2d 769 (2006) (excluding Simon Cole's testimony, reasoning, "what Dr. Cole cannot do in challenging the admissibility of the State's fingerprint evidence, he equally cannot do here in purportedly challenging the weight of said evidence"); *United States v. Taylor*, No. CR 07-1244 WJ, 2009 WL 6338569 (D.N.M., Sept. 30, 2009) (firearms identification case where the testimony of Adina Schwartz was excluded at trial because "to allow Dr. Schwartz to give this testimony during a trial would not be, as Defendant argues, to allow a healthy 'battle of the experts.' Instead it would be to engage in what this Court considers the highly questionable practice of allowing a second *Daubert* hearing to play out in front of the jury.") *Daubert* itself says, "Vigorous cross examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence." *Daubert*, 509 U.S. 579 (1993). If the best argument for admitting this evidence notwithstanding its shakiness is that understanding the weaknesses is not beyond the ken of the jury, it is imperative that the defense be permitted the "traditional and appropriate means" of attacking this evidence. *Id.*

¹²⁶ 188 F. Supp. 2d 549 (E.D. Penn. 2002).

discriminating measure of competence. But he nonetheless found that the absence of any evidence showing a substantial error rate provided some evidence that the error rate was tolerably low.¹²⁷ His bottom line was, more or less, that fingerprint evidence wasn't exemplary, but it was good enough: "[T]o postpone present in-court utilization of this 'bedrock forensic identifier' pending such research would be to make the best the enemy of the good."¹²⁸ All things considered, it's valid (enough).

C. *Dividing the Baby: The Solomonic Compromise*

The third approach, to which I will give the most attention because it represents the most thoughtful judicial approach taken to date to assessing the admissibility of forensic science evidence in court, is what I am calling the "Solomonic compromise." This approach reflects judges' efforts to split the difference between admissibility and exclusion in a way that superficially seems to make sense, but becomes, I will suggest, increasingly problematic upon careful reflection. This approach permits the expert to testify about similarities and differences in the patterns at issue, but prohibits or limits the expert from reaching expert conclusions about the *meaning* of those similarities. *Llera Plaza I*—the opinion Judge Pollak vacated after a preliminary hearing—took precisely this approach. His explanation in his original opinion is worth quoting in detail:

Since the court finds that ACE-V does not meet *Daubert's* testing, peer review, and standards criteria, and that information as to ACE V's rate of error is in limbo, the expected conclusion would be that the government should be precluded from presenting any fingerprint testimony. But that conclusion—apparently putting at naught a century of judicial acquiescence in fingerprint identification processes—would be unwarrantably heavy-handed. The *Daubert* difficulty with the ACE-V process is by no means total. The difficulty comes into play at the stage at which, as experienced fingerprint specialists Ashbaugh and Meagher themselves acknowledge, the ACE-V process becomes "subjective"—namely, the evaluation stage. By contrast, the antecedent analysis and comparison stages are, according to the testimony, "objective": analysis of the rolled and latent prints and comparison of what the examiner has observed in

¹²⁷ *Id.* at 566.

¹²⁸ *Id.* at 572. For a thoughtful critical analysis of this opinion, see generally David H. Kaye, *The Nonscience of Fingerprinting: United States v. Llera-Plaza*, 21 QUINNIPIAC L. REV. 1073 (2003).

the two prints. Up to the evaluation stage, the ACE-V fingerprint examiner's testimony is descriptive, not judgmental. Accordingly, this court will permit the government to present testimony by fingerprint examiners who, suitably qualified as "expert" examiners by virtue of training and experience, may (1) describe how the rolled and latent fingerprints at issue in this case were obtained, (2) identify and place before the jury the fingerprints and such magnifications thereof as may be required to show minute details, and (3) point out observed similarities (and differences) between any latent print and any rolled print the government contends are attributable to the same person. What such expert witnesses will not be permitted to do is to present "evaluation" testimony as to their "opinion" (Rule 702) that a particular latent print is in fact the print of a particular person. The defendants will be permitted to present their own fingerprint experts to counter the government's fingerprint testimony, but defense experts will also be precluded from presenting "evaluation" testimony. Government counsel and defense counsel will, in closing arguments, be free to argue to the jury that, on the basis of the jury's observation of a particular latent print and a particular rolled print, the jury may find the existence, or the non-existence, of a match between the prints.¹²⁹

There are several aspects to note regarding *Llera Plaza I*'s logic. First, the court clearly and decisively recognizes that the latent fingerprint identification evidence did not fare well under *Daubert*'s strictures.¹³⁰ But then, because some part of the evidence is "descriptive, not judgmental" and in partial deference to the long judicial acceptance of this form of proof, the judge determined that outright exclusion would be "unwarrantably heavy-handed."¹³¹ His compromise was to permit the expert to show the jury the similarities in the prints at issue, to point their attention toward the data that the expert would typically use to derive his or her conclusion, but to prohibit the expert from actually providing that conclusion or opinion to the jury. The idea is that the expert would merely be showing "objective" data to the jury; it would then be up to the jury, without overt expert assistance, to decide what

¹²⁹ *Llera Plaza (I)*, 179 F. Supp. 2d at 516. Note that while Judge Pollak was the first judge to take this compromise approach to the admissibility of latent fingerprint evidence in particular, he was following an approach taken by several other courts in the context of expert evidence on handwriting identification evidence. The first judge to take this approach was Judge Matsch, in the trial of Timothy McVeigh for the Oklahoma City bombing. No written opinion was issued, but the transcript of the discussion of this issue is available at Pre-Trial Transcr., *United States v. McVeigh*, 1997 WL 47724 (D. Colo. Feb. 5, 1997). Judge Nancy Gertner was the first judge to issue a written opinion in support of this approach in *United States v. Hines*, 55 F. Supp. 2d 62 (D. Mass. 1999). See *infra* notes 137-138 and accompanying text.

¹³⁰ *Llera Plaza (I)*, 179 F. Supp. 2d at 516.

¹³¹ *Id.*

meaning to give to the data, and what conclusion, if any, to reach about the likelihood that the fingerprints it had been shown derived from a common source.

While other courts have not yet taken this approach to fingerprint evidence (and of course Pollak himself reversed course in *Llera Plaza II*), a number of other judges have adopted this same structural compromise—description without ultimate conclusion—in cases involving other kinds of pattern identification evidence, notably handwriting and firearms identification.¹³²

At first glance, this compromise approach seems to be a clever and appropriate strategy.¹³³ Merely pointing out the similarities and differences in two visible patterns without providing any conclusion does, it seems to me, reduce or even eliminate the *Daubert* problems with the evidence. To be sure, it is not altogether clear that an expert's testimony is even necessary to point out similarities and differences in two visual exemplars, as the jury members can look at the images for themselves and thus have access to the same visual data with or without the expert's testimony. However, due to his or her training and experience, an expert may well be better at *seeing* those similarities and differences. Lay jurors may therefore be meaningfully assisted in their own observations and examinations of the visual stimuli by having the expert point out precisely what is worth looking at, how to look, and how to see minutiae, and both the similarities and differences in the visual exemplars, for themselves.

The experts, in other words, may, under this approach, provide the factfinder with a kind of educative expertise—the expert does not ask for deference to his or her authority, but rather, teaches the jury members how to see the patterns present in the fingerprint or bullet or handwriting sample for

¹³² See, e.g., *United States v. Hines*, 55 F. Supp. 2d 62 (D. Mass. 1999) (involving handwriting identification); *United States v. Green*, 405 F. Supp. 2d 104 (D. Mass. 2005) (involving firearms identification); *United States v. Glynn*, 578 F. Supp. 2d 567 (S.D.N.Y. 2008) (permitting a firearms identification to be deemed “more likely than not” but not permitting individualization in absolute terms). There are several other handwriting identification cases that have followed *Hines* in taking this approach. See generally *Risinger, Cases Involving*, *supra* note 8.

¹³³ Indeed, several academic commentators have noted it with approval. See, e.g., Michael J. Saks, *Protecting Factfinders From Being Overly Misled, While Still Admitting Weakly Supported Forensic Science into Evidence*, 43 TULSA L. REV. 609 (2007); Robert P. Mosteller, *Finding the Golden Mean with Daubert: An Elusive, Perhaps Impossible, Goal*, 52 VILL. L. REV. 723, 760-62 (2007).

themselves.¹³⁴ It is as if the expert is teaching the jury how to read music, instead of playing the notes on the piano himself. The expert provides a lesson in “how to see,” and the jury then exercises its own vision and reaches its own conclusion. Framed in this way, it is fair to say that the expert assistance in pointing out similarities and any difference meets the “helpfulness” requirement for expert testimony under *Daubert*.¹³⁵

Furthermore, the lack of knowledge about error rates, the non-existence of objective standards for determining a match, and the lack of statistical models for determining the probabilities of a match do not seem nearly as problematic under *Daubert*, at least at first glance, when no *conclusion* about the meaning of the match is being provided. As Judge Gertner explained in a case involving handwriting identification,

[The expert's] account of what is similar or not similar in the handwriting of [the defendant] and the robber can be understood and evaluated by the jury. . . . This is not rocket science, or higher math. Her conclusion of authorship, however, has a difference resonance: “Out of all of my experience, and training, I am saying that he is the one, the very author.” That leap may not at all be justified by the underlying data; and in the context of this case, is extraordinarily prejudicial.¹³⁶

If the expert confines herself to pointing out similarities and differences to the factfinder, while eschewing all conclusions, then the questions regarding the legitimate strength of the expert's conclusions, or the lack of a statistical model to justify a claim about the probabilities associated with a match, or the appropriate error rates associated with the expert's conclusions, all become moot. No match, and, indeed, no ultimate conclusion regarding identification or its absence, is being introduced into evidence.

Perhaps, then, this compromise provides an exemplary way to navigate away from the awkward spot in which the forensic sciences now find themselves? Certainly, the small handful of judges that have put forward this compromise view are to be commended, for they are wrestling valiantly with a

¹³⁴ Ronald J. Allen & Joseph S. Miller, *The Common Law Theory of Experts: Deference or Education?*, 87 NW. U. L. REV. 1131 (1993). For an extensive discussion of education versus deference in the context of handwriting identification experts, see generally Mnookin, *Scripting Expertise*, *supra* note 1.

¹³⁵ *Daubert*, 509 U.S. 579 at 591.

¹³⁶ *Hines*, 55 F. Supp. 2d at 69.

set of hard questions about how to handle these powerful but inadequately tested forms of proof. It may well be that the Solomonic compromise is the best practical alternative we can come up with at present if courts deem exclusion too draconian a remedy. This compromise eliminates the excessively strong and presently unjustified claims about the strength of forensic experts' conclusions. It represents an effort to acknowledge, on the one hand, the likely power of this evidence, while, on the other hand, not treating the evidence as if it passes *Daubert* with flying colors like *Havvard*.

And yet, just like King Solomon's proposal to divide the baby between the two women claiming to be its mother, this compromise approach is, unfortunately, fundamentally unsatisfying. Why so? Upon careful analysis, there are two significant difficulties that arise. First, it may be substantially more difficult, as a practical matter, to eliminate the evaluative aspects from testimony than one might expect. For example, Judge Gertner wrote in *Hines*, "The witness can be cross examined, as she was, about why this difference was not considered consequential, while this difference was, and the jury can draw their own conclusions."¹³⁷ But notice what the witness would be testifying to here. While she would not be testifying to an ultimate conclusion about authorship, she *would* be testifying as to why some differences are considered consequential and others are not, rather than simply presenting the visual content of the handwriting exemplars and pointing out similarities and/or differences. This is, already, evaluative rather than merely descriptive.¹³⁸ She would be saying, more or less, "this kind of difference is still consistent with the two samples coming from the same person's writing, while this other difference is not something we would expect to see in two samples of writing from the same person." Perhaps the expert's inferences regarding the interpretive

¹³⁷ *Id.*

¹³⁸ Moreover, as I pointed out some time ago, taken to the extreme, this would result in little more than a semantic difference between this "restricted" approach and permitting conclusion testimony under *Daubert*. If an expert can say to the jury, "Look at these seventeen meaningful similarities, and here's why they are really meaningful" and, "Look at these three apparent differences, and here's why I don't think they count for anything," the expert's own conclusion about authorship or identification would be completely clear to the jury, even if the expert did not actually use the words "match" anywhere in his or her testimony. See generally Mnookin, *Scripting Expertise*, *supra* note 1; accord, Risinger, *Cases Involving*, *supra* note 8, at 510; D. Michael Risinger & Jeffrey Loop, *Three Card Monte, Monty Hall, Modus Operandi and "Offender Profiling"*, 24 CARDOZO L. REV. 193, 209-10 (2002).

meaning of this data are correct, and perhaps not. The point is that either way, this is *not* merely showing the jury the existence of similarities and differences but is, already, evaluating their meaning, determining which differences truly “count” in favor of a conclusion of authorship and which may not. And thus we need again to ask a set of already familiar questions: What data shows that the expert can do what she or he claims to be able to do? Is there data to support the notion that a certain kind of difference is consistent with authorship and another kind is not? How accurate are such conclusions? What is their error rate?

We have returned, in essence, straight back to the *Daubert* problem, albeit having drilled down the analysis to a more particularized, local, level, rather than asking validity questions with respect to an overall conclusion about authorship of the writing. Similarly, in the case of fingerprint identification, consider the expert’s belief that a particular apparent visual difference between the two prints is an artifact rather than an actual discrepancy. If the expert is permitted to testify that he or she believes that a particular difference is not a true discrepancy, that too raises all the same questions about the extent to which data supports the ability of an expert to accurately distinguish between source discrepancies and differences that are not interpretively meaningful, “real” differences in the source prints versus those differences that derive from the process of taking and making the images. At present, adequate published data regarding the reliability and scientific validity of these mid-level inferences is largely non-existent.

This analysis might suggest a straightforward answer: perhaps what is needed is simply to take a more complete end-run around *Daubert* by carefully prohibiting what we might call “intermediate” inferences about the evidence. Judges could permit experts truly to testify only to what they see, literally just to point the factfinder toward details on the bullet, the print, or the handwriting exemplar that the layperson might not have noticed without expert assistance. If experts truly limited themselves only to description,¹³⁹ wouldn’t that

¹³⁹ I am here ignoring the question of whether there is actually any such thing as description without interpretation—my own view would be that there is not. Nonetheless, these are matters of degree, and prohibiting all inferences about the likely significance of the minutiae observed does get closer to being “just” description than does permitting the expert’s overtly evaluative statements.

eliminate the lurking validity problems associated with mid-level evaluations and conclusions? Leave it entirely to the jury to decide whether particular visual differences ought or ought not to be taken as meaningful discrepancies in two fingerprints. Leave it to the jury to decide whether the quantity and types of differences in two handwriting samples suggest distinct authorship or remain consistent with one person having some degree of inevitable variation in how they form their letters.

If we truly limited the expert to “pure” description, this might indeed solve the lurking *Daubert* issues, though it would do so by radically curtailing the expert’s role. It should be noted that this would also be quite an unusual approach to expertise—part of what is special about experts’ roles under the Federal Rules of Evidence is precisely that experts are typically given significantly more leeway than lay witnesses to provide their opinions and conclusions.¹⁴⁰ Placing forensic experts into a kind of Lockean straitjacket—permitting them only to testify to what they can empirically observe, rather than allowing them to share the *inferences and judgments* they make about what they see—is, in a sense, to de-authorize them as experts. Much of what a forensic expert thinks of as his or her expertise—which is precisely the experience-based ability to assess and analyze the image; to differentiate signal from noise and artifact from discrepancy; and to evaluate whether two patterns did or did not come from the same source—would no longer be permitted. This is not necessarily a bad outcome—indeed, it might well spur an increased interest among forensic scientists themselves in promoting and participating in the research that would permit the courts to grant them a more significant evaluative role. But it is important to recognize that if we take this compromise seriously, the permitted testimony must be quite limited, significantly more curtailed than it was, for example, in *Hines* itself.¹⁴¹

Moreover, we should recognize that juries will likely still find such extremely curtailed evidence quite probative. Juries walk in the door with prior, culturally-based views about evidence and what kinds of evidence count, and most jurors probably arrive with a deeply-held belief in the significant power of pattern identification evidence, just like my seatmate

¹⁴⁰ See generally FED. R. EVID. 701-02.

¹⁴¹ *Hines*, 55 F. Supp. 2d 62.

on the airplane whose beliefs about forensic science derived mostly from her television-viewing. So even if all an expert did was point out similarities, with absolutely no testimony whatsoever about the *meaning* of these similarities, a jury might be quite prepared to *presume* that those similarities imply that the two impressions come from a common source.

This leads to the second, more serious difficulty. The still-larger problem with the Solomonic compromise is that if experts cannot provide some data-based, research-justified evidence supporting their conclusions, it is not clear that the impression evidence should be admissible at all. The problem, in a sense, migrates: it stops being a problem relating to the *adequacy of the evidence supporting scientific validity*, and becomes a problem of *relevance and probative value*.

To put the point in terms of the Federal Rules of Evidence, radically curtailing the experts' testimony solves the Rule 702 expert evidence problem, but at the cost of creating a Rule 401/403 problem. Rule 401 requires that all evidence be relevant, and the Federal Rules define as relevant that evidence which makes "the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."¹⁴² Rule 403 permits the exclusion of relevant evidence when "its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury."¹⁴³

The problem, in a nutshell, is that if the expert cannot provide the factfinder with any admissible evidence about the *meaning* of the visual similarities and differences, then the factfinder has no rational basis for assessing the probative value of these observations. Certainly the factfinder can look at the visual stimuli herself, but what legitimate basis does she have for making an inference about probative value given whatever quantum of similarities she observes? The key question about the meaning of any pattern identification evidence is how much support it provides for the inference that two patterns do or do not share a common source. How often would we expect to see any given degree of similarity from two bullets that did *not* come from the same gun, or from two

¹⁴² FED. R. EVID. 401.

¹⁴³ FED. R. EVID. 403.

fingerprints that were not actually impressions from the same finger?

To answer this question, juries could draw on two potential sources of knowledge: information provided by the evidence presented to them at trial via an expert, or their own experience. But in the Solomonic compromise, we are prohibiting the first kind of information, because it does not at present meet *Daubert's* strictures. So juries are left entirely to their own resources and devices—only their own experience can help them assess the probative value of the patterns at issue. But a serious problem arises because—with the possible exception of handwriting identification—juries simply do not have any meaningful experience on which to draw for these conclusions. Non-experts are not in the habit of looking closely at ridge minutiae on fingerprints to develop intuitions about how much similarity might exist on the tips of two different individuals' fingers. Ordinary people do not encounter bullet striations in their regular life. Jurors therefore have literally no personal, experience-based information that would provide any rational basis for evaluating the similarities and differences that the expert helped them to notice. While the evidence might nonetheless squeak by Rule 401, given the very low threshold for defining relevance under the Rules of Evidence, the lack of any rational basis for assessing its meaning makes the evidence both prejudicial and potentially misleading, thus rendering it excludable under Rule 403.¹⁴⁴ One can think of the issue like this: because we believe forensic expert testimony often has probative value, it is therefore relevant—but because we really do not have any rational way to assess that probative value, its admissibility seriously risks being both misleading and prejudicial.

To be sure, we regularly permit jurors to engage in lay assessments of “frequencies” and to determine for themselves probative value in the trial process. If a victim who saw the perpetrator testifies that the person who assaulted her had grey hair, a beard, and a tattoo of a purple parrot on his forearm, and the defendant also has those characteristics, we

¹⁴⁴ Certainly it is a reasonable inference to say that two prints or two bullets that have many similarities to one another are more likely to come from a common source than two that do not. For this reason, the evidence does, I suppose, meet Federal Rule of Evidence 401's low standard, which deems relevant that evidence which has “any tendency” to make a matter more or less probative standard. The problem that we know it has some tendency, but cannot presently quantify how much or how little power it really has.

do not require any data to be presented to the jury on the frequency with which any or all of these characteristics actually exist within the relevant suspect population. Nor do we require any formalized model for assessing the likelihood that these traits would co-exist within a given individual within that population, or whether the different characteristics tend to be statistically independent from one another or not.¹⁴⁵ We let the jury decide for itself precisely how much power and how much probative value to give to the evidence introduced that shows that the perpetrator and the defendants share certain characteristics in common. The fact that individual juror's subjective assessments of the frequency of purple parrot tattoos might be wildly off-base does not render the evidence inadmissible.

Why, then, is the evidence of the similarities in bullet striations or ridge detail on fingerprints any different from the evidence of the purple parrot tattoo? Yes, juror assessments of the evidentiary power of the pattern similarities in a fingerprint might be substantially inaccurate, but that risk exists in the tattoo scenario as well. The difference, I would suggest, is that we believe that jurors' ordinary lives provide them with some legitimate basis—albeit partial and imperfect—for assessing the frequency of purple parrot tattoos on people's forearms. In ordinary life, people see each other's forearms—and each other's tattoos—with some regularity. Most of us probably do not go around counting how many people have tattoos, or keeping track of how many birds we have ever seen tattooed on other people's bodies. But nonetheless, our individual paths navigating through the crowded world do give us an experiential basis for having a rough empirical sense of just how rare or common purple parrot tattoos might be.¹⁴⁶ Note, in a sense, that our individual

¹⁴⁵ Indeed, in the famous case of *People v. Collins*, 438 P.2d 33 (Cal. 1968), a case that came to stand for the reluctance of courts to frame questions in overtly probabilistic terms, the effort to use the “product” rule to figure out a combined probability of a variety of specific characteristics—ranging from the fact that the couple was interracial to the fact that she had blonde hair and a ponytail and he had a beard—was strongly criticized. For the classic critique of excessive confidence in probabilistic thinking in court, see generally Lawrence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971). However, one of the (numerous) problems in *Collins* was that the numbers presented to the jurors were purely speculative—they were simply the expert's illustration of possible numbers to illustrate how the product rule worked, rather than numbers based on empirical reality. *Collins*, 438 P.2d 33.

¹⁴⁶ There is a stronger argument for taking the “compromise” approach in handwriting identification cases, precisely because our ordinary lives do give us some

relationships to the evaluation of parrot tattoos are rather like the forensic expert's relationship to the evaluation of patterns—they are experience-based, non-quantitative, and without any known error rate. The difference however, is in role: we expect juries to bring their common knowledge to their role as factfinder and to their assessment of the evidence, and we do not require this common knowledge to be validated or scientific. The same cannot be said for the expert required to meet the strictures of *Daubert*.

In addition, the strong cultural belief in forensic science evidence, based first and foremost on its roughly one hundred years of courtroom use, likely means that even without any actual experience of their own looking at comparable patterns, jurors will believe they *do* know the meaning of the similarities they are taught by the expert to observe. With latent fingerprint examination, the tradition of experts individualizing in court—without adequate data or research to support their claim—may well mean that when a jury is presented with latent fingerprint identification evidence without any expert conclusions, it will nonetheless believe itself quite capable of drawing a conclusion of identity. But this confidence and belief does not derive from the first-hand experience of jury members looking with care at fingerprints or bullet striations. Rather, it would stem from their prior belief that latent fingerprints can individualize, and hence if these prints appear reasonably similar, then they probably did come from the same person. They may not ever have looked at fingerprints for themselves, but from courtroom dramas on television, from criminal cases they have read about in the newspaper, from CSI, from their myriad cultural experiences in the world, they may well believe not only that everyone's fingerprints are different, but also that two similar prints must necessarily have come from the same finger.¹⁴⁷

meaningful experience looking at people's handwriting. Non-experts do have some experience looking at writing, attempting to gauge authorship, and have an experience-based sense of what degree of similarity of writing one might expect to see in the writing of the same author versus two different authors. To be sure, experts may be better than non-experts at assessing the meaning of these similarities and differences—but the point here is not who is better at this, or by what degree, but rather, whether non-experts have a lay basis for evaluation sufficient to justify admissibility of the similarities if there is no admissible evidence about their meaning. To put it differently, handwriting is a lot more like the purple parrot tattoo.

¹⁴⁷ The cultural mythology of fingerprints remains deep and widespread; even my seven-year-old son told me the other day that everyone's fingerprints are different.

Notice the concerning kind of feedback loop occurring here. The substance of the excluded evidence—conclusions about the common source of the latent print and the exemplar—is likely to be *presumed* by the jury precisely because that now-excluded evidence has generally been admitted in the past. If my hunch about likely jury expectations and prior understanding is correct, then the common knowledge that will be doing a good deal of interpretive work for the jury when they try to assess the fingerprint is *the very evidence that was excluded*. That certainly raises a significant Rule 403 problem.

Therefore, unless the expert can provide the jury with some valuable, data-driven basis for interpreting the *meaning* of the similarities and differences that are presented, the pattern identification evidence presented under the Solomonic compromise still ought to raise significant admissibility concerns. This is not because the experts' descriptions are insufficiently reliable under Rule 702, but because of the lack of any basis—apart from the jury's likely prior belief in the very conclusions that have been excluded—for meaningfully evaluating the probative value of the evidence.

The Solomonic compromise is thus far more problematic than it initially appears. However, I recognize that it might nonetheless be a reasonable second-best solution given the present impasse, and I therefore do believe it has a legitimate place in the judicial “toolkit” for assessing pattern identification evidence. Moreover, there might be ways, at trial, to dislodge, at least partially, whatever prior beliefs the jurors had about the basis for individualization or the meaning of a “match.” If the Solomonic compromise were coupled with effective evidence to demonstrate that there is not a statistical basis for reaching a conclusion about whether the two patterns come from a common source, nor any validated metrics for evaluating how much similarity is needed to warrant such a conclusion, the jury might call into question its prior assumption that similarity necessarily meant identity.

There is little doubt that evidence of similarity of bullet striation patterns or handwriting similarities or friction-ridge patterns on fingerprints often does have probative value. The question is how much. The hard question facing judges is what to do given that the answer to that “how much” question is, “we really don't quite know.” It seems clear that on the basis of current empirical knowledge, judges should absolutely not permit conclusions of individualization to be made by experts,

and softening an individualization conclusion by framing it as “opinion” rather than “fact” does not enhance either its validity or its research basis. Nor should either judges or experts kid themselves that adding in some fudge words like “to a reasonable certainty” changes the analysis at all.¹⁴⁸

The harder question is whether the fact that we don’t know precisely how much probative value to assign to evidence of any given quantum of similarity—and that a jury assessing the meaning of these similarities likely has no first-hand life experience to help assess probative value—ought to lead to the evidence’s exclusion. As a matter of logic, without some meaningful basis for assessing the probative value, the evidence becomes literally uninterpretable—and this should rationally argue in favor of exclusion. But given that experts’ experience and our collective cultural experience over the last century with fingerprint identification evidence supports the inference that the probative value of this evidence is likely to be quite substantial, the Solomonic compromise, notwithstanding its awkwardness, might be as good an option as any for the moment. The Solomonic compromise, in a sense, is an approach based on a bet that the cultural belief in forensic science will, when more research has been conducted, largely prove to have been warranted. The more confident a judge is that future research will be likely to validate the claims fingerprint examiners have been making in court for the last 100 years, the less troubled she might be by letting those conclusions in through the back door, via jurors’ prior beliefs about forensic science and its credibility. Put like that, the Solomonic compromise becomes a more nuanced version of judicial ACE-V—all things considered, it’s valid enough to permit, so long as we force it into an uncomfortable straitjacket, a straitjacket that cannot be expected to stay on within the jury room.

Thus, even though it does have an appropriate role in the judicial toolkit, the Solomonic compromise is, at best, an awkward doctrinal solution. If only we had the data to support them, it would be far better to permit expert inferences about the *meaning* of similarities and *conclusions* about probative value. It is also worth noting that outright exclusion would put significantly more pressure on the forensic science community to cooperate with and to lobby for the research that would

¹⁴⁸ *Taylor*, 663 F. Supp. 2d at 1179.

bring them back into the evidentiary fold. Nonetheless, I can certainly understand the attraction of the Solomonic compromise as a second-best solution for probative but under-researched evidence that has a long history of legal acceptance.

IV. IN DEFENSE OF EXCLUSION (FOR NOW)

The other viable alternative is, obviously, outright exclusion, at least in some circumstances.¹⁴⁹ Given judges' treatment of forensic science evidence so far, arguing for exclusion of this evidence can feel, from a practical standpoint, more than a little quixotic.

And yet, the truth of the matter is that at 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 sloppiness¹⁵⁰—exclusion would be a great motivator for pursuing the research necessary to justify admissibility.

Moreover, I want to suggest that the use of the exclusion option—which I do think should have a central place in the judicial toolkit, given the present lack of an adequate research basis supporting validity—could also be quite short-term in most pattern identification arenas. For as I will suggest, what ought to be the minimum necessary information to establish adequate validity under *Daubert* is simply not that onerous. 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

¹⁴⁹ I do not explicitly address the question of when courts should exclude, and when the Solomonic approach is warranted. Partly that is because I think, in many instances, either approach could be a legitimate exercise of judicial discretion given the current state of our knowledge. Moreover, the choice ought to be informed by the particulars—were I a judge ruling on admissibility, I would be far more likely to exclude a single AFIS-generated match than evidence that linked multiple, high-quality prints to several different fingers of the same individual. Even without formal metrics for sufficiency, common sense—and *Kumho Tire*—tell us that all identification tasks are not created equal, and the judicial response can, and sometimes should, vary as a result.

¹⁵⁰ See generally JAY D. ARONSON, *GENETIC WITNESS: SCIENCE, LAW AND CONTROVERSY IN THE MAKING OF DNA PROFILING* (2007); DAVID H. KAYE, *THE DOUBLE HELIX AND THE LAW OF EVIDENCE* (2009); Jennifer L. Mnookin, *People v. Castro*, in *EVIDENCE STORIES* (Richard Lempert, ed. 2006).

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 particular case is sufficiently close.

More generally, as I have also argued elsewhere, I want to suggest that under *Daubert* it is the ‘testing’ criteria that should matter most.¹⁵¹ I want to make a distinction between explanation or description, on the one hand, and testing on the other. Explanation of the methods and descriptions of the processes used by an expert should not be permitted to substitute for adequate testing of validity. It is this mistake, I believe, that has often plagued the courts when evaluating forensic science. Judges hear about the ACE-V process and they listen to examiners describing their approach, and judges are persuaded, it seems, that this methodology therefore works.¹⁵² They, like the factfinders, see the similarities magnified and put up on a giant chart, and they “see” the method in action, and believe its power. Even putting aside that ACE-V is not, in fact, much of a specified methodology, the more important point is that the courts should care less about the details of the method at issue or its seeming plausibility, and more about what *evidence* there is to support the conclusion that the methods actually work.

In other words, judges have been lulled by plausible descriptions and seemingly persuasive explanations of forensic science techniques into dismissing the importance of the nearly complete lack of empirical support for the experts’ claims. All of the opinions discussed above as examples of legal “ACE-V” have this quality. These judges recognized the lack of testing, but found enough within the description of the method that seemed credible that they decided that the technique passed muster.¹⁵³ This, in my view, is a mistake.

To be sure, the pattern identification techniques of forensic science do have a certain intuitive plausibility to them, and their early acceptance was linked, in no small part, to this cultural plausibility.¹⁵⁴ We all have some experience identifying handwriting for example: no doubt most of us believe that we

¹⁵¹ See generally Jennifer L. Mnookin, *Of Black Boxes, Instruments, and Experts: Testing the Validity of Forensic Science*, 5 EPISTEME 343 (2008) [hereinafter Mnookin, *Of Black Boxes*].

¹⁵² See *supra* Part III.B.

¹⁵³ See generally *supra* notes 105-113 and accompanying text.

¹⁵⁴ See generally Mnookin, *Fingerprint Evidence*, *supra* note 1; Mnookin, *Scripting Expertise*, *supra* note 1.

could likely distinguish our mother's handwriting from that of our closest friend. We can stare down at our fingerprints as well, and can even see for ourselves how the ridges and whorls on our fingers vary from those of others. Moreover, these patterns have a particular quality—they are *semi-legible* in that they can be seen and their differences can be noted by all of us, not just by experts, while at the same time, they still do require significant expert analysis and interpretation.

But the inherent plausibility and the semi-legible quality of these materials combine to make it particularly easy for judges to be seduced by description and explanation into failing to ask what they ought to be asking both as a matter of doctrine and as a matter of logic. Their focus should be on the degree of empirical support for the actual, specific claims being made. They should be asking precisely what evidence supports this particular evidentiary claim. With latent fingerprint evidence, for example, the most central question ought to be: how accurate are examiners when matching latent prints to a particular source; latent prints which are often partial, frequently smudged, and perhaps even distorted? For any of the forensic sciences, what judges ought to ask under *Daubert* is precisely this: what empirical support shows that the expert can actually do what she claims to do? What data, what testing, would be necessary to justify the claims being made in the expert's testimony?

An important corollary of this focus on testing is that it ought to be permissible under *Daubert* for the technique or method to be a kind of "black box."¹⁵⁵ A "black box" is a technique or method that we do not necessarily understand, but that we can nonetheless test to see what it does and how it works. If there is sufficient testing to show us *that* it works, I do not believe that the proffering party should be required under *Daubert* to show *how* it works. My argument is that it is far less important to pry open this black box than it is to ask whether the technique has been tested under conditions similar to those at issue in court. Peering inside the black box to see *how* it works is less critical for an assessment of validity than assessing whether input/output testing shows *that* it works.¹⁵⁶ Indeed, peering inside, on its own, ought to be deemed neither necessary, nor sufficient under *Daubert*.

¹⁵⁵ Mnookin, *Of Black Boxes*, *supra* note 151.

¹⁵⁶ *Id.* On black box testing, see Risinger, *A Glass Nine-Tenths Full*, *supra* note 72, at 31 n.66 and Risinger & Saks, *Science and Nonscience in the Courts*, *supra*

To make the point more concretely, when it comes to fingerprint evidence, for example, what we really need in order to justify admissibility is, at a minimum, some very good proficiency tests to show us what experts can do and to gain information about how often they make mistakes. These proficiency tests ought to be appropriately difficult and should mirror the range of difficulty found in actual casework. In addition, and critically, they should include some of what fingerprint experts would call “tough idents”—prints that are particularly difficult to identify.¹⁵⁷ These proficiency tests should ideally be part of the normal stream of casework, so the examiner doesn’t know she is being tested and therefore possibly perform her analysis with a greater degree of care than usual.¹⁵⁸

Recall that one of the research lacunae with respect to fingerprint evidence is that we do not yet have an operational statistical model of fingerprints, a model that could provide us with empirically grounded information about the likelihood that two prints selected at random would both have a specified set of minutiae. Recall also that ACE-V is extremely vague, and does not come close to providing a fully developed and adequately articulated method with detailed specifications.¹⁵⁹ Recall, in addition, that interpretation of fingerprints is subjective, without shared norms or rules about what is required. My argument is that these gaps in our knowledge base, though unfortunate, are not fatal. More precisely, I want to suggest that neither a detailed specification of method, nor statistical validation of frequencies ought to be seen as necessary criteria for using fingerprints as evidence in court under *Daubert’s* strictures.

In other words, it ought to be acceptable for latent print examiners themselves to operate as a kind of black box—for the

note 1, at 40-43. To be sure, I am largely skating around hard questions about whether testing can ever be wholly “black box,” or what some degree of internal understanding of the method is necessary in order to design good black box tests, and if so, how much. Still, the key point is to urge courts to focus not on asking experts to *explain what they do*, but rather to *show that what they do works*.

¹⁵⁷ There is not presently any validated metric of difficulty for fingerprint comparisons. Nonetheless, examiners do have at least informal understandings of what makes particular comparisons more or less difficult.

¹⁵⁸ There is presently a “black box” study of more or less this sort underway, being conducted by Austin Hickland through the FBI. It sounds, from the descriptions, like an extremely promising project. However, at this point neither the research protocol nor the results have been made public.

¹⁵⁹ Haber & Haber, *supra* note 1.

examiners themselves to be the instrument, the technique. If we know through proficiency tests that they get the right answers a very high proportion of the time over a range of circumstances that mimic what they encounter in actual cases, their testimony ought to be admissible under *Daubert*. Even if we do not fully know how their method operates—indeed, even if *they* do not fully understand it themselves—this ought not to prevent us from making evidentiary use of their conclusions if we have sufficient information about how accurate they are and what circumstances and conditions seem to increase the risk of error.

Let me make one thing clear. I am certainly not opposed to research and inquiry that peers inside the black box to learn more about how the methodology works and aims to improve it. There are some impressive efforts underway to model fingerprint evidence statistically, and this is extremely important research that I fully support.¹⁶⁰ Of course, fingerprint experts should continue to hone their methods, continue to work for better understanding of how they could improve their processes, and if researchers can help practicing forensic scientists to develop validated standards for interpretation to increase objectivity, that will be all to the good. All of these efforts at better understanding, formalizing methods, and improving practices—forms of opening up the black box and peering inside—ought to be welcomed, celebrated, and encouraged, and perhaps most importantly, funded.

They just ought not to be necessary requirements for admissibility in court. Nor, standing alone, should they be sufficient, unless they are also accompanied by evidence that these methods, in actual practice rather than in theory, truly work. To make the argument by analogy: Which would make a driver more comfortable—knowing that there was substantial theoretical knowledge suggesting that the brakes on her vehicle *ought* to work, because of a great deal of scientific study of the theoretical mechanisms underlying the brakes—or a substantial quantity of actual testing of identical vehicles showing that the brakes *have* worked consistently in conditions that mirror her situation? While both forms of knowledge are valuable, I would posit that the actual, on-the-ground testing, quite appropriately gives us a good deal more comfort than the theoretical knowledge standing alone.

¹⁶⁰ Neumann et al, *supra* note 31.

I am arguing therefore, that the courts' central focus, especially for a kind of evidence that does not presently have a formalized method, should be on testing. Courts should also insist upon a close relationship between the testing that has been done and the claims that an expert makes. This is what *Daubert* calls the question of "fit" and it is also part of what *Kumho Tire* emphasizes as the need to focus on "the task at hand."¹⁶¹ Strong claims should require strong tests to back them up. For example, fingerprint examiners claim to be able to individualize—to match a print to one unique source, to one finger on one person out of everyone on earth who has ever lived or who will ever live.¹⁶² There is, as I have already stated, not yet sufficient validation to support that claim. Nor do I think that proficiency tests of the sort that I am describing and suggesting as a prerequisite to admissibility would be sufficient to support this claimed ability to individualize. Even with excellent proficiency tests, experts would therefore need to modulate and moderate the strength of their conclusions to some degree. But excellent performance on difficult proficiency tests might be sufficient to support a conclusion of a softer sort, a conclusion such as, "Based on my knowledge, testing and experience, I would not expect to see this degree of similarity between two prints unless they came from a common source." This is still, in a way, fudging—for without a working statistical model, we cannot have a robust and quantified sense of the likelihood two prints with a given degree of similarity might have come from different sources. But if we knew that examiners in general—and *that* examiner in particular—had succeeded in proficiency tests that required her to make difficult identifications and equally difficult exclusions, we could legitimately believe that her knowledge and experience did adequately support her opinion that they came from a common source.

A signal advantage of focusing on testing rather than explanation is that it has the virtue of being manageable, from a practical point of view. There are simply no insurmountable obstacles to developing appropriately difficult proficiency tests. The roadblocks to doing it thus far have been cultural and institutional, not scientific. By contrast, developing a valid statistical model of fingerprint evidence is a daunting task.

¹⁶¹ *Daubert*, 509 U.S. 579 at 591.

¹⁶² See *supra* note 35 and accompanying text.

Even with significant research, it likely is years away, perhaps even decades. By contrast, proficiency test development could be done quite rapidly. And I have little doubt that if courts began to exclude fingerprint evidence for the lack of such tests, the requisite testing mechanisms would be developed in extremely short order.

Implicit in my argument here for a focus on testing is a quasi-“best evidence” approach to the evaluation of expert testimony.¹⁶³ I do not believe that *Daubert* should be understood as requiring a fixed and unchanging amount of evidence in any particular area. Rather, the question ought to be whether the proffered expert evidence is *as reliable as it can reasonably be, considering the context and circumstances*. Validity under *Daubert* should not be understood as an on/off switch, or as an all-or-nothing proposition, in which items of evidence are inherently reliable or unreliable. The question for the court is whether they are reliable *enough*—and this depends both on what inferences are sought to be drawn from them, and partly on whether the evidence offered was as reliable as possible under the circumstances. To put it differently, to pass muster under *Daubert*, the judge must have some legitimate justification for believing that the evidence is sufficiently reliable that a jury should hear it. And the evidence of validity should be as strong as it reasonably can be, given the circumstances. It is, therefore, partly because proficiency tests are genuinely ‘do-able’ that courts ought to require them. Concomitantly, courts should hesitate before finding that studies or research that are beyond current scientific capacity are nonetheless required under *Daubert*.

In other words, if an expert—or an entire field—has done as much as can reasonably be done to establish validity, and our still-imperfect information suggests significant probative value, excluding the evidence because our validity knowledge is incomplete is not likely to be justice-enhancing—especially if our knowledge includes reasonable estimates of error rate, so that the factfinder can, at least in theory, adjust its assessment of probative value accordingly. But by the same token, if an expert—or a field—fails to undertake those tasks which could reasonably be done to establish validity, and

¹⁶³ See generally Mnookin, *Of Black Boxes*, *supra* note 151; see also Edward Cheng, *Mitochondrial DNA: Emerging Legal Issues*, 13 J.L. & POL’Y 99 (2005); Dale Nance, *Reliability and the Admissibility of Experts*, 34 SETON HALL L. REV. 191 (2003); Dale Nance, *The Best Evidence Principle*, 73 IOWA L. REV. 227 (1988).

instead simply tells the court, “Trust me!” the court should balk at the lack of data. Under this approach, evidentiary assessment will necessarily be dynamic. A reasonable amount of research at Time A may not continue to be adequate at Time B, as time passes, and the potential for developing relevant data on the salient questions increases.

It is worth noting that what drives me to a ‘better evidence’ principle is very much the same kinds of concerns that have motivated Margaret Berger in her important work on causation in tort law. Professor Berger has been acutely concerned about the difficulties plaintiffs have faced in establishing causation in toxic torts cases.¹⁶⁴ Throughout her career, she has been sensitive to the significant difficulties that scientific uncertainty poses for our legal system. In the torts context, she has, therefore, searched for ways to protect plaintiffs’ interests while simultaneously respecting the need for high-quality scientific information within our system of adjudication. Professor Berger has offered some insightful and creative solutions to the tension between the need for good scientific information, and the need to recognize that *Daubert* (at least as interpreted by courts in the pharmaceutical torts context) often expects too much. She has therefore searched for principled ways around general causation, sometimes through the possibility of expanding other kinds of tort claims—like the right to informed choice in making a decision about medication, or the duty of a company to keep itself informed of risks through the pursuit of reasonable research.¹⁶⁵

By contrast, in the forensic science challenges, courts have interpreted *Daubert* as to expect rather too little, instead of too much. But the underlying concerns—how do we deal with uncertain knowledge, and how can we generate the right incentives within the legal system both to do justice and to produce better information—are strikingly similar. A ‘better evidence’ principle may be our best bet—certainly in the forensic science context, but perhaps more generally across the board as well, though its potential application in other expert contexts goes entirely beyond the scope of this article.

¹⁶⁴ See, e.g., Margaret Berger, *Eliminating General Causation: Notes Towards a New Theory of Justice and Toxic Torts*, 97 COLUM. L. REV. 2117 (1997); Margaret Berger, *Upsetting The Balance Between Adverse Interests: The Impact of the Supreme Court’s Trilogy on Expert Testimony in Toxic Tort Litigation*, 64 LAW & CONTEMP. PROBS. 289 (2001).

¹⁶⁵ See, e.g., Margaret Berger & Aaron Twerski, *Uncertainty and Informed Choice: Unmasking Daubert*, 104 MICH. L. REV. 257 (2006).

Neither a focus on testing nor a more general 'best evidence' approach to expert evidence is a panacea. There will still be difficult questions for courts, especially regarding just how much testing is required to justify admissibility. I do not pretend that answering those questions will be easy or straightforward. But they are, I think the right questions for courts to ask. They are no doubt challenging questions for many judges, most of whom typically lack much background in science, and who perhaps went to law school in part precisely to stay far away from such technical matters. Judges typically therefore will lack epistemic competence—the ability to evaluate the knowledge being offered the way an insider to the field would evaluate it.¹⁶⁶

Nonetheless, I believe that by focusing on testing, courts are less likely to be misled, less likely to be lulled by compelling-sounding but as-of-yet unproven explanations, and less likely to gloss over remarkable gaps in what is known. Testing is a narrower, more tightly bounded inquiry. It explicitly directs judges to the question of what data is available to support the expert's conclusions, and to focus on the specific fit between the data and the claim. Might judges misunderstand a study, fail to notice methodological flaws, or misconstrue what inferences can legitimately be drawn from a given research result? Absolutely. But at least these errors of interpretation would be the result of a focus on the data itself—which suggests, first and foremost, that there actually *is* some data on the relevant questions. Given our present state of affairs, this in itself would be a major improvement.

In addition, recall that I said at the outset of this article that we needed a two pronged solution to our problems of forensic science—both stricter scrutiny by the courts *and* greater federal support for research and regulatory oversight. Part of why we absolutely need the two-pronged solution is as a check on these concerns about the epistemic competence of the courts. If there is a research establishment funding work on forensic science—including the funding needed both to develop proficiency tests, and to evaluate them—it is highly likely that academic researchers will increasingly be attracted to the study of forensic science. Research follows money like bees

¹⁶⁶ See Jennifer L. Mnookin, *Expert Evidence, Partisanship and Epistemic Confidence*, 73 BROOK. L. REV. 1009 (2008); see also Scott Brewer, *Expert Testimony and Intellectual Due Process*, 107 YALE L. REV. 1535 (1998); Michael Pardo, *The Field of Evidence and The Field of Knowledge*, 24 L. & PHIL. 321 (2005).

follow pollen. Fund it and they will come. And the structures of academic science and social science—with peer review, academic evaluation, as well as the review and evaluation that could be offered by a new federal agency if one ever were to be developed—could all offer a useful check on the quality of the proficiency tests that need to be developed.

There is one final advantage that would result from judges taking *Daubert* seriously in the context of pattern identification, and, therefore, either dramatically restricting or excluding evidence. At present, virtually no one believes that a new federal agency—NIFS, as suggested in the NAS report is likely. Whether the federal government will even create a new institutional space somewhere within existing federal agencies, dedicated to important issues of forensic science, is far from certain. There is a serious risk that at the agency level, nothing of genuine or transformative import will result from the NAS report.

However, if judges were to begin to take seriously the implicit lessons of the report, the odds would change dramatically. Why so? Because it would mean that a number of important forensic sciences would be likely to find themselves—only temporarily, I would expect—excluded from the courtroom due to a lack of adequate testing of the validity of their results and methods.

And if that were to begin to happen, what would ensue? My strong instinct is that the degree of interest in NIFS or at least a more modest surrogate for NIFS within an existing agency, would skyrocket. Obviously, the reason for judges to apply *Daubert* meaningfully and therefore likely exclude at least some forensic science evidence cannot primarily be to incentivize the creation of a federal agency. Judges' appropriate focus is on the particular cases before them, and explicit reference to the external incentive effects their decisions might generate is not a first priority; even thinking in such terms may make some judges uneasy. But in this instance, there is a happy convergence: to wit, if the courts begin to do their job well, it is likely to help bring about precisely the broader public policy initiatives that are also necessary in this area.

CONCLUSION

What will the future hold for forensic science? If we look back on these methods and techniques twenty years from now, what will we see? It is perhaps entirely safe to predict that no

matter what happens, the practices of forensic science still won't look like the television version, so my airplane seatmate is destined for ongoing disappointment with the real world. But I do not think it is utterly unrealistic to hope that the pattern identification sciences will be on a much more substantial and solid empirical footing than they are today. Ironically, they may well look somewhat less strong than they do at present. Testifying experts will certainly no longer be able to espouse an error rate of zero, and they will likely need to give up the claim that they are able to individualize. But by acknowledging their weaknesses, and honestly assessing their capacities and limitations, they will truly be far stronger than they are at present, and far more worthy of credence and respect. Will forensic science transform itself as it should? That will depend, I believe, in significant part on judges, and whether they are prepared, at long last, to evaluate pattern identification evidence with their eyes wide open and their heads out of the sand.

For all of our sake, I hope they are up to the challenge.