

# TAB 4: PROPOSALS TO CHANGE FRE 702

# **RULEMAKING POSSIBILITIES: EFFORTS OF THE U.S. JUDICIAL CONFERENCE ADVISORY COMMITTEE ON EVIDENCE RULES TO ADDRESS THE RECENT CHALLENGES TO FORENSIC EXPERT TESTIMONY**

**Daniel J. Capra**

Reed Professor of Law

Fordham Law School

Reporter to the Advisory Committee on Evidence Rules

David E. Bernstein & Eric G. Lasker, *Defending Daubert: It's Time to Amend Federal Rule of Evidence 702*, 57 Wm. & Mary L. Rev. 4 (2015).

The 2000 Advisory Committee on the Federal Rules of Evidence has not accomplished its goal of requiring courts to adopt a more rigorous and structured approach to scrutiny of expert testimony. The 2000 Advisory Committee amended FRE 702. It now states that trial courts must examine the factual foundation of expert testimony, the reliability of the expert's methodology, and the expert's application of her methodology to the facts at issue. Although courts always apply FRE 702, there are divisions in how it is applied, and the divisions are the same divisions that existed before the 2000 amendments. Bernstein and Lasker propose a new amendment to Rule 702.

Part I of Bernstein and Lasker's article surveys the history of the 2000 amendments to Rule 702. Part II describes the resistance of some courts to adopting the requirements of Amended Rule 702. Part III surveys case law subsequent to Rule 702's amendments. Part IV proposes specific revisions to Rule 702 to remedy lingering problems.

Bernstein and Lasker recommend the following amendments to Rule 702:

“A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise ***if the testimony satisfies each of the following requirements:***

- (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on ~~sufficient~~ facts or data ***that reliably support the expert's opinion;***
- (c) the testimony is the product of reliable ***and objectively reasonable*** principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case ***and reached his conclusions without resort to unsupported speculation.***

***Appeals of district court decisions under this Rule are considered under the abuse-of-discretion standard. Such decisions are evaluated with the same level of rigor regardless of whether the district court admitted or excluded the testimony in question.***

***This Rule supersedes any preexisting precedent that conflicts with any or all sections of this Rule.***”

\*added text is bolded and italicized. Removed text has been struck through.

Naftali A. Schachtman, Esq., PC

- [News](#)
- |
- [Publications & Presentations](#)
- |
- [Biographical](#)
- |
- [Blog](#)
- |
- [File Library](#)

« [Woodside on Access to Underlying Research Data  
Vaccine Court Inoculated Against Pathological Science](#) »

## On Amending Rule 702 of the Federal Rules of Evidence

No serious observer or scholar of the law of evidence can deny that the lower federal courts have applied *Daubert* and its progeny, and the revised Federal Rule of Evidence 702, inconstantly and inconsistently, in their decisions to admit or exclude proffered expert witness opinion testimony. Opponents of trial court “gatekeeping” of expert witnesses applaud the lapses in hopes that the gates have been unhinged and that there will be “open admissions” for expert witness testimony. These opponents latch on to the suggestion that the Rules favor “liberal” admissibility with a libertine; they lose sight of the meaning of “liberal” that conveys enlightened, with an openness to progress and salutary change, and the claims of knowledge over blind faith. Supporters of gatekeeping lament the courts’ inability or unwillingness to apply a clear statutory mandate that is designed to improve and ensure the correctness of fact finding in the federal courts. A few have decried the lawlessness of the courts’ evasions and refusals to apply Rule 702’s requirements.

Given the clear body of Supreme Court precedent, and the statutory revision to Rule 702, which was clearly designed to embrace, embody, enhance, and clarify the high Court precedent, I did not think that an amendment to Rule 702 was needed to improve the sorry state of lower court decisions. Professor David Bernstein and lawyer Eric Lasker, however, have made a powerful case for amendment as a way of awakening and galvanizing federal judges to their responsibilities under the law. David E. Bernstein & Eric G. Lasker, “[Defending \*Daubert\*: It’s Time to Amend Federal Rule of Evidence 702](#),” 57 *William & Mary L. Rev.* 1 (2015) [cited below as Bernstein & Lasker].

Bernstein and Lasker remind us that Rule 702 is a statute<sup>[1]</sup> that superseded inconsistent prior judicial pronouncements. The authors review many of the more egregious cases that ignore the actual text of Rule 702, while adverting to judicial gloss on the superseded rule, and even to judicial precedent and dicta pre-dating the *Daubert* case itself. Like the Papa Bear in the Berenstain Bear family, the authors show us how not to interpret a statute properly, through examples from federal court decisions.

### The Dodgers’ Dodges

***Questions about whether expert witnesses properly applied a methodology to the facts of a case are for the jury, and not the proper subject of gatekeeping.***



As Bernstein and Lasker document, this thought- and Rule-avoidance dodge is particularly shocking given that the Supreme Court clearly directed close and careful analysis of the specific application of general principles to the facts of a case.<sup>[2]</sup> Shortly after the Supreme Court decided *Daubert*, the Third Circuit decided a highly influential decision in which it articulated the need for courts to review every step in expert witnesses' reasoning for reliability. *In re Paoli RR Yard PCB Litig.*, 35 F.3d 717, 745 (3d Cir. 1994). The Paoli case thus represents the antithesis of a judicial approach that asks only the 10,000 foot level question whether the right methodology was used; Paoli calls for a close, careful analysis of the application of a proper methodology to every step in the case. *Id.* ("any step that renders the analysis unreliable ... renders the expert's testimony inadmissible ... whether the step completely changes a reliable methodology or merely misapplies that methodology").

While the *Paoli* approach is unpopular with some judges who might prefer not to work so hard, the Advisory Committee heartily endorsed *Paoli's* "any step" approach in its Note to the 2000 amendment. Bernstein & Lasker at 32. Bernstein and Lasker further point out that the Committee's Report, Professor Dan Capra, acknowledged, shortly after the amendment went into effect, that the *Paoli* "any step" approach had a "profound impact" on the drafting of amended Rule 702. Bernstein & Lasker at 28.<sup>[3]</sup>

Having demonstrated the reasons, the process, and the substance of the judicial and legislative history of the revised Rule 702, Bernstein and Lasker are understandably incensed at the lawlessness of circuit and trial courts that have eschewed the statute, have ignored Supreme Court precedent, and have retreated to vague judicial pronouncements that trace back to before some or any of the important changes occurred in Rule 702.<sup>[4]</sup>

### ***Let's Cherry Pick and Weigh of the Evidence; Forget the Scale***

Along with some courts' insistence that trial judges may not examine the application of methods to the facts of a case, other courts, perhaps mindful of their citation practices, have endorsed "cherry picking" as a satisfactory methodology for partial expert witnesses to support their opinions. *Id.* at 35-36. Our law review authors also trace the influence of plaintiffs' counsel, through their "walking around money" from the breast implant litigation, in sponsoring anti-*Daubert*, anti-gatekeeping conferences, at which prominent plaintiffs' advocates and expert witnesses, such as Carl Cranor presented in favor of a vague "weight of the evidence" (WOE) approach to decision making. *Id.* at 39. Following these conferences, some courts have managed to embrace WOE, which is usually packaged as an abandonment of scientific standards of validity and sufficiency, in favor of selective review and subjective decisions. To do this, however, courts have had to ignore both Supreme Court precedent and the clear language of Rule 702. In *Joiner*, the high Court rejected WOE, over the dissent of a single justice,<sup>[5]</sup> but some of the inferior federal courts have embraced the dissent to the exclusion of the majority's clear holding, as well as the incorporation of that holding into the revised Rule 702.<sup>[6]</sup> An interesting case of judicial disregard.

### **Other Dodges**

The law review authors did not purport to provide an exhaustive catalogue of avoidance and evasion techniques. Here is one that is not discussed: shifting the burden of proof on admissibility to the opponent of the expert witness's opinion:

"Testimony from an expert is presumed to be helpful unless it concerns matters within the everyday knowledge and experience of a lay juror."

*Earp v. Novartis Pharms.*, No. 5:11–CV–680–D, 2013 WL 4854488, at \*3 (Sept. 11, 2013). See also *Kopf v. Skyrn*, 993 F.2d 374, 377 (4th Cir.1993); *accord Koger v. Norfolk S. Ry. Co.*, No. 1:08–0909, 2010 WL 692842, at \*1 (S.D.W.Va. Feb. 23, 2010) (unpublished).

Whence comes this presumption? Perhaps it is no more than a requirement for the opponent to object and articulate the flaws before the trial court will act. But the “presumption” sure looks like a covert shifting of the burden of proof for the requisite reliability of an expert witness’s opinion, which burden clearly falls on the proponent of the testimony.

### The Proposed Amended Rule 702

There are several possible responses to the problem of the judiciary’s infidelity to basic principles, precedent, and legislative directive. Bernstein and Lasker advance amendments to the current Rule 702, as a blunt reminder that the times and the law have changed, really. Here is their proposed revision, with new language italicized, and deleted language shown to be struck:

“Rule 702. Testimony by Expert Witnesses

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

- (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on ~~sufficient~~ facts or data *that reliably support the expert’s opinion*;
- (c) the testimony is the product of reliable *and objectively reasonable* principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case *and reached his conclusions without resort to unsupported speculation*.

*Appeals of district court decisions under this Rule are considered under the abuse-of-discretion standard. Such decisions are evaluated with the same level of rigor regardless of whether the district court admitted or excluded the testimony in question. This Rule supersedes any preexisting precedent that conflicts with any or all sections of this Rule.”*

Bernstein & Lasker at 44-45.

Before discussing and debating the changes, we should ask, “why change a fairly good statute just because lower courts evade its terms?” The corrupt efforts of SKAPP<sup>[7]</sup> to influence public and judicial policy, as well as the wildly one-sided *Milward* symposium,<sup>[8]</sup> which the authors discuss, should serve as a potent reminder that there would be many voices in the review and revision process, both from within plaintiffs’ bar, and from those sympathetic to the litigation industry’s goals and desires. Opening up the language of Rule 702 to revision could result in reactionary change, driven by the tort bar’s and allies’ lobbying. The result could be the evisceration of Rule 702, as it now stands. This danger requires a further exploration of alternatives to the proposed amendment.

Rule 702 has had the benefit of evolutionary change and development, which have made it better and also possibly burdened with vestigial language. To be sure, the rule is a difficult statute to draft, and while the authors give us a helpful start, there is many problems to be subdued before a truly workable working draft can be put be forward.

The first sentence's new language, "*the testimony satisfies each of the following requirements,*" is probably already satisfied the use of "and" between the following numbered paragraphs. Given the judicial resistance to Rule 702, the additional verbiage could be helpful; it should be unnecessary. The conditionality of "if," however, leaves the meaning of the Rule unclear when that condition is not satisfied. The Rule clearly signifies that "if" in the introductory sentence means "only if," and the law and litigants would be better off if the Rule said what it means.

### **Proposed Subsection (b)**

(b) the testimony is based on ~~sufficient~~ facts or data *that reliably support the expert's opinion*;

The authors do not make much of a case for striking "sufficient." There will be times when there are perfectly good facts and data supporting an expert witness's opinion, in a completely reliable opinion, but the supporting facts and data do not support an epistemic claim of "knowledge," because the support is indeterminate between the claim and many other competing hypotheses that might explain the outcome at issue. The reliably supporting facts and data may amount to little more than a scientific peppercorn, and really be too much of too little to support the claim. Deleting "sufficient" from subsection b could be a serious retrograde move, which will confuse the judiciary more than instruct it.

The revised subsection also fails to address the integrity of the facts and data, and the validity of how the data were generated. To be sure, Rule 703 could pick up some of the slack, but Rule 703 is often ignored, and even when invoked, that rule has its own drafting and interpretation problems. See "[Giving Rule 703 the Cold Shoulder](#)" (May 12, 2012); "[RULE OF EVIDENCE 703 — Problem Child of Article VII](#)" (Sept. 19, 2011). Also missing is an acknowledgment that the facts or data must often be analyzed in some way, whether by statistical tests or some other means. And finally, there is the problem in that reliable does not necessarily connote valid or accurate. Subsection (b) thus seems to cry out for additional qualification, such as:

"the testimony is based on sufficient facts or data, reliably, accurately, and validly ascertained, and analyzed, which facts or data reliably and validly support the expert's opinion"

### **Proposed Subsection (c)**

Bernstein and Lasker propose modifying this subsection to inject "and objectively reasonable" before "principles and methods." The authors do not explain what objectively reasonable principles and methods encompass, and they qualification does seem self-explanatory. Perhaps they are calling for principles and methods that are "generally accepted," and otherwise justified as warranted to produce accurate, true results? Is so, that might be a helpful addition.

### **Proposed Subsection (d)**

Here the authors bolster the language of the subsection with a prohibition against using unsupported speculation. OK; but would supported or inspired or ingenious speculation be any better? Subsection

(a) speaks of knowledge, and it should be obvious that the expert witness's opinion has an epistemic warrant to be something more than a mere subjective opinion.

Whether Bernstein and Lasker have opened a can or a Concordat of Worms remains to be seen.

---

[1] The authors provide a great resource on the legislative history of attempts to revise 702, up to and including the 2000 revision. The 2000 revision began with a proposed amendment from the Advisory Committee in April 1999. The Standing Committee on Rules of Practice and Procedure approved the proposal, and forwarded the proposed amendment to the Judicial Conference, which approved the amendment without change in September 1999. The Supreme Court ordered the amendment in April 2000, and submitted the revised rule to Congress. Order Amending the Federal Rules of Evidence, 529 U.S. 1189, 1195 (2000). The revised Rule 702 became effective on December 1, 2000. *See also* Bernstein & Lasker at 19 n. 99 (citing Edward J. Imwinkelried, "Response, Whether the Federal Rules of Evidence Should Be Conceived as a Perpetual Index Code: Blindness Is Worse than Myopia," 40 *Wm. & Mary L. Rev.* 1595, 1595-98 (1999) (noting and supporting the Supreme Court's interpretation and application of the Federal Rules of Evidence as a statute, and subject to the judicial constraints on statutory construction). For a strident student's pro-plaintiff view of the same legislative history, *see* Nancy S. Farrell, "[\*Congressional Action to Amend Federal Rule of Evidence 702: A Mischievous Attempt to Codify Daubert v. Merrell Dow Pharmaceuticals, Inc.\*](#)," 13 *J. Contemp. Health L. & Pol'y* 523 (1997).

[2] *General Electric Co. v. Joiner*, 522 U.S. 136 (1997) (reviewing and analyzing individual studies' internal and external validity, and rejecting plaintiffs' argument that only the appropriateness of the methodology in the abstract was subject of gatekeeping); *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 156-57 (1999) ("stressing that district courts must scrutinize whether the principles and methods employed by an expert have been properly applied to the facts of the case") (quoting what was then the proposed advisory committee's note to Rule 702, Preliminary Draft of Proposed Amendments to the Federal Rules of Civil Procedure and Evidence: Request for Comment, 181 F.R.D. 18, 148 (1998)).

[3] citing Stephen A. Saltzburg, Edwin J. Imwinkelried, & Daniel J. Capra, "Keeping the Reformist Spirit Alive in Evidence Law," 149 *U. Pa. L. Rev.* 1277, 1289-90 (2001). The authors note that other circuits have embraced the Paoli "any steps" approach. Bernstein & Lasker at 28 at n. 152 (citing *Paz v. Brush Engineered Materials, Inc.*, 555 F.3d 383, 387-91 (5th Cir. 2009); *McClain v. Metabolife Int'l, Inc.*, 401 F.3d 1233, 1245 (11th Cir. 2005); *Dodge v. Cotter Corp.*, 328 F.3d 1212, 1222 (10th Cir. 2003); *Amorgianos v. Nat'l R.R. Passenger Corp.*, 303 F.3d 256, 267 (2d Cir. 2002) (quoting *In re Paoli*, 35 F.3d at 746).

[4] *See, e.g., City of Pomona v. SQM N. Am. Corp.*, 750 F.3d 1036, 1047 (9th Cir. 2014) (rejecting the Paoli any step approach without careful analysis of the statute, the advisory committee note, or Supreme Court decisions); *Manpower, Inc. v. Ins. Co. of Pa.*, 732 F.3d 796, 808 (7th Cir. 2013) ("[t]he reliability of data and assumptions used in applying a methodology is tested by the adversarial process and determined by the jury; the court's role is generally limited to assessing the reliability of the methodology – the framework – of the expert's analysis"); *Bonner v. ISP Techs., Inc.*, 259 F.3d 924, 929 (8th Cir. 2001) ("the factual basis of an expert opinion goes to the credibility of the testimony, not the admissibility, and it is up to the opposing party to examine the factual basis for the opinion in cross-examination").

[5] *General Electric Co. v. Joiner*, 522 U.S. 136, 146-47 (1997) (holding that district court had the “discretion to conclude that the studies upon which the experts relied were not sufficient, whether individually or in combination, to support their conclusions that Joiner’s exposure to PCB’s contributed to his cancer”). Other federal and state courts have followed *Joiner*. See *Allen v. Pa. Eng’g Corp.*, 102 F.3d 194, 198 (5th Cir. 1996) (“We are also unpersuaded that the ‘weight of the evidence’ methodology these experts use is scientifically acceptable for demonstrating a medical link between Allen’s EtO exposure and brain cancer.”). For similar rejections of vague claims that weak evidence add up to more than the sum of its parts, see *Hollander v. Sandoz Pharm. Corp.*, 289 F.3d 1193, 1216 n.21 (10th Cir. 2002); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 608 (D.N.J. 2002); *Caraker v. Sandoz Pharm. Corp.*, 188 F. Supp. 2d 1026, 1040 (S.D. Ill. 2001); *Siharath v. Sandoz Pharm. Corp.*, 131 F. Supp. 2d 1347, 1371 (N.D. Ga. 2001), *aff’d sub nom. Rider v. Sandoz Pharm. Corp.*, 295 F.3d 1194 (11th Cir. 2002); *Merck & Co. v. Garza*, 347 S.W.3d 256, 268 (Tex. 2011); *Estate of George v. Vt. League of Cities & Towns*, 993 A.2d 367, 379-80 (Vt. 2010).

[6] *Milward v. Acuity Specialty Products Group, Inc.*, 639 F.3d 11, 17-18 (1st Cir. 2011) (reversing the exclusion of expert witnesses who embraced WOE). *Milward* has garnered some limited support in a few courts, as noted by Bernstein and Lasker; see *In re Fosamax (Alendronate Sodium) Prods. Liab. Litig.*, Nos. 11-5304, 08-08, 2013 WL 1558690, at \*4 (D.N.J. Apr. 10, 2013); *Harris v. CSX Transp., Inc.*, 753 S.E.2d 275, 287-89, 301-02 (W. Va. 2013).

[7] “[SKAPP A LOT](#)” (April 30, 2010).

[8] “[Milward Symposium Organized by Plaintiffs’ Counsel and Witnesses](#)” (Feb. 13, 2013); [<http://perma.cc/PW2V-X7TK>].

This entry was posted on Saturday, October 17th, 2015 at 4:29 am and is filed under [Expert Witnesses](#), [Rule 702](#), [Scientific Evidence](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. Both comments and pings are currently closed.

Comments are closed.



[Web Development by eMaak](#)

© Nathan A. Schachtman, Esq., P.C.,  
and Nathan A. Schachtman, all written content.

[Disclaimer Contact](#)

325 E. 79th Street, Ste. 16-D  
New York, NY 10075  
[nathan@schachtmanlaw.com](mailto:nathan@schachtmanlaw.com) 212.600.4912



Geoffrey Stewart Morrison & William C. Thompson, *Assessing the Admissibility of a New Generation of Forensic Science Voice Comparison Testimony*, 18 Colum. Sci. & Tech. L. Rev. 326 (2017).

Sometimes, disputes arise in criminal cases over the identity of a speaker on an audio recording. To resolve the dispute, a forensic voice comparison may be performed. Forensic voice comparison evidence was last used in court in 2003, and thrown out by the judge. Since then there have been no reported instances where this type of evidence has overcome a *Daubert* challenge. However, there have been significant advances in forensic voice comparison technology in the last 15-20 years. This article offers a framework for assessing the strength of forensic voice comparison testimony.

First, Morrison and Thompson provide a “primer” on voice comparison technology. Second, they describe the frameworks that practitioners use when and reporting the strength of the evidence. They describe seven different frameworks. The leading framework is the “Likelihood-Ratio Framework,” which requires the practitioner to consider: 1) the probability of obtaining the observed properties of the voice on the questioned-speaker recording if it were produced by the known speaker; and 2) the probability of obtaining the observed properties of the voice on the questioned-speaker recording if it were produced not by the known speaker, but by some other speaker from the relevant population.

Next, Morrison and Thompson discuss how to test for and control validity and reliability. They also discuss contextual bias, and recommend minimizing contextual bias by avoiding exposure of practitioners to “task-irrelevant” information (information that is not necessary for assessing the strength of forensic evidence). They also mention several ways to shield practitioners from task-irrelevant information: context management procedures, having a case manager as a middleman between analysts and clients, and sequential unmasking.

Morrison and Thompson also discuss admissibility and apply *Daubert* to forensic voice comparison testimony. They also review and critique the use of forensic voice technology in the recent case *U.S. v. Ahmed*. In conclusion, Morrison and Thompson outline how they believe a forensic voice comparison would have to be conducted to produce testimony that is admissible under *Daubert* (I have attached a copy of their proposal).

- The practitioner conducted one automatic analysis (Comparison 1), got a result, then modified the system (by adding an imposter set), reran the analysis, and got a result more favorable to the prosecution than the first result. A forensic practitioner should avoid acting in a way that could give the impression that they are cherry picking results, *i.e.*, that they tested multiple systems and then selected the one which was most favorable to the party instructing them.<sup>2 4</sup>

### C. Conclusion with Respect to the *Ahmed* Testimony

The forensic practitioner in *Ahmed* used a mixture of approaches: auditory, acoustic-phonetic, and automatic. The results of all of the analyses were subjective judgments. Even for the automatic subsystem, which calculated likelihood ratios using quantitative measurements and statistical models, the practitioner did not directly report the calculated values, but instead used them as inputs to making a subjective decision. The way the results from \*413 each analysis were combined was also a subjective judgment. In general the procedures were not transparent, and were not described in sufficient detail that they could be replicated by another suitably qualified practitioner.

With respect to the *Daubert* factors, the practitioner did not empirically test the validity and reliability of his system under conditions reflecting those of the case under investigation. There is no evidence that he followed any standards which we would consider indicators of trustworthiness. Although there were some peer-reviewed publications supporting some aspects of his approach, their relevance for assessing the trustworthiness of his overall conclusions was limited. Whether his approach could be considered generally accepted in the relevant scientific community is unclear. Indeed, whether *any* particular approach is generally accepted at this time is unclear. While his approach may be in line with common practice among practitioners, it is not in line with current practice in the scientific research community. Clearly, we believe that the testimony did not satisfy the *Daubert* criteria and should not have been admitted.

Shortly after the hearing, the prosecution offered what the defense viewed as a favorable plea bargain and the case was resolved with a negotiated plea, rendering the admissibility issue moot. Although some might interpret this development as evidence that the prosecution feared losing the *Daubert* hearing and the case, there is no way to know how the court would have ruled. It remains to be seen how courts will view forensic voice comparison evidence when it is offered in future cases.

## VIII. MEETING THE *DAUBERT* STANDARD: WHAT WOULD A POTENTIALLY ADMISSIBLE FORENSIC VOICE COMPARISON ANALYSIS LOOK LIKE?

Our critique of the testimony presented in *Ahmed* has been overwhelmingly negative. This does not, however, mean that we believe that forensic voice comparison testimony could never be found admissible under *Daubert*. We think that, in practice, only approaches based on relevant data, quantitative measurement, and statistical models would be able to satisfy the *Daubert* criteria. Below we outline how we believe a forensic voice comparison \*414 would have to be conducted in order to produce testimony which could potentially be found admissible under *Daubert*.<sup>2 5</sup>

1. To facilitate transparency and replicability, the forensic practitioner should document in their report or in bench notes all decisions they make and all actions they take. All parties should be made aware of the existence of these notes, and they should be provided to all parties upon request. All substantial decisions and actions should also be documented in the forensic report. On the basis of the report, bench notes, and a copy of the practitioner's standard operating procedures and other appendices, another suitably qualified forensic practitioner (or researcher) should be able to critique the first practitioner's decisions and actions



and potentially replicate what the first forensic practitioner did. If anything is unclear in the report and appendices, the second practitioner should be able to find the answer in the first practitioner's notes. The second forensic practitioner should not have to guess what the first forensic practitioner actually did.

2. To reduce the potential for contextual bias, the practitioner should take steps to avoid being exposed to task-irrelevant information, *i.e.*, information about the case which is not necessary for them to perform their forensic voice comparison analysis. In large laboratories, a case manager may be \*415 assigned to handle communication with the client and other parties, and only pass on to the practitioner task relevant information. In smaller laboratories the practitioner should ask the client up front to not provide task-irrelevant information.

3. Based on an examination of the questioned-speaker recording, and relevant information provided by the client and other parties as may be appropriate given the circumstances of the case, the practitioner should formulate the details of the same-speaker hypothesis and the different-speaker hypotheses that they plan to assess. The different-speaker hypothesis must include the definition of the relevant population. Before proceeding, the suitability of these hypotheses should be confirmed with the client and other parties as may be appropriate given the circumstances of the case. The hypotheses, including the relevant population, should be clearly described in the report.

4. Based on an examination of the known- and questioned-speaker recordings, and relevant information provided by the client and other parties as may be appropriate given the circumstances of the case, the practitioner should describe what they understand to be the speaking styles and recording conditions of the known-speaker recording and the questioned-speaker recording. All reasonable enquiries should be made to obtain technical details about recording systems, *etc.* These conditions should be clearly described in the report.

5. If the practitioner believes that *a priori* the conditions of the recordings are so poor that the performance of their forensic voice comparison system will be so poor that it is unlikely to be of value to the court, they should inform the client of this before proceeding. The client may still request that the practitioner proceed, but this will be an informed decision. If the client decides not to have the practitioner proceed with a \*416 particular comparison, this should be documented in the report, and no further analyses should be conducted on the relevant recordings.

6. The known- and questioned-speaker recordings should be prepared by selecting only portions of the recordings which actually contain speech of the speaker of interest. Interlocutor speech, transient noises, and stretches of silence or background noise should be excluded from the analysis. (This will reveal one aspect of the recording conditions, the net durations of the known-speaker and the questioned-speaker speech.)

7. The practitioner should obtain a sample of voice recordings representative of the relevant population and reflecting the speaking styles and recording conditions of the known-speaker recording and the questioned-speaker recording. The sample may come from an existing database, or new data may need to be collected.

The practitioner must be satisfied that the sample recordings are sufficiently representative and reflective of the relevant population, speaking styles, and recording conditions. The report must explain how the forensic practitioner sampled the speakers, and how they replicated or simulated the conditions. Sufficient detail must be provided so that the judge at an admissibility hearing has a basis on which to consider whether the recordings are sufficiently representative and reflective. We would expect the opposing parties to seek expert advice in this topic, and debate the merits before the judge during an admissibility hearing (if the testimony is admitted, this topic may also be argued before the trier-of-fact in relation to weight).

8. The relevant population sample recordings should be prepared in the same manner as the known- and questioned-speaker recordings.

9. The practitioner should split their data into at least two <sup>2 6</sup> separate parts: a training set and a test set. Statistical models should not be trained and tested on the same data.

10. To reduce the potential for contextual bias, the practitioner should use a system based on relevant data, quantitative measurements (*e.g.*, measurements of acoustic properties of the voice recordings), and statistical models. The output of the statistical model should be directly reported, it should not be used as input to a subsequent subjective judgment process.

11. The system should be trained and optimized using the training data, which reflect the relevant population, speaking styles, and recording conditions of the case. Ideally, a second forensic practitioner should check the first forensic practitioner's work at this stage in search of any potential mistakes. Once the forensic practitioner is satisfied with the training and optimization of the system, the system should be frozen, *i.e.*, no subsequent changes to the system will be allowed.<sup>2 7</sup>

12. The practitioner should then use the test data to empirically assess the performance of their system. The system as a whole should be tested, including any components depending on the particular human operator. The system which is tested should be the same system which will actually be used to compare the known- and questioned-speaker recordings. The results of the tests should be documented in the report, and an explanation of how to interpret any numeric or graphical results should be provided in the report or in an appendix. Sufficient detail <sup>2 8</sup> \*418 should be provided to assist the judge at an admissibility hearing to decide if system performance is sufficient to warrant admission of the testimony (if the testimony is admitted, this question may also be argued before the trier-of-fact in relation to weight). Ideally, a second forensic practitioner should check the first forensic practitioner's work at this stage in search of any potential mistakes. Once the tests have been conducted, they should not be repeated in search of better results. The system should not be altered and then retested on the same data set.

13. The last step in the analysis should be to actually compare the known- and questioned-speaker recordings. The numeric likelihood ratio generated by the system should be reported as the strength of evidence statement. The report, or an appendix, should include an explanation of the likelihood ratio

framework so that the judge at an admissibility hearing and the trier-of-fact at trial can understand how to appropriately interpret the result. Once the likelihood ratio for the comparison of the known- and questioned-speaker recordings has been obtained, the system should not be altered or retested, and the likelihood ratio should not be recalculated in search of a better answer.<sup>29</sup>

Such procedures would, we believe, be potentially admissible under *Daubert* because they are logically correct, robust to cognitive bias, transparent and replicable, and include demonstration of degree of validity and reliability under conditions reflecting those of the case under investigation. If the judge at an admissibility hearing is satisfied (1) that the test data are sufficiently representative of the relevant population and sufficiently reflective<sup>419</sup> of the speaking styles and recording conditions of the known-speaker recording and the questioned-speaker recording, and (2) that the empirically demonstrated degree of validity and reliability of the system under these conditions is adequate, then the system will have passed what we consider to be the most important *Daubert* criterion, *i.e.*, “whether the reasoning or methodology underlying the testimony is scientifically valid and ... whether that reasoning or methodology properly can be applied to the facts in issue,” including “whether it can be (and has been) [empirically] tested,” and “in the case of a particular scientific technique ... consider[ation of] the known ... rate of error.”<sup>220</sup>

## IX. CONCLUSION

We have argued that the most important *Daubert* criterion for deciding the admissibility of an implementation of any approach to forensic voice comparison (be it auditory, acoustic-phonetic non-statistical, acoustic-phonetic statistical, or automatic) is whether it has been empirically tested under conditions reflecting those of the particular case under investigation, and found to be sufficiently valid and reliable. We see this as the direct primary indicator of scientific validity, and the other *Daubert* criteria as secondary proxy indicators. If the judge accepts that the test data are sufficiently representative of the relevant population and sufficiently reflective of the conditions of the case under investigation, they can then consider whether the empirically demonstrated performance of the system under those conditions is sufficient to warrant admission. We have also argued that, because of the substantial case-to-case variability in relevant population, speaking styles, and recording conditions, system performance will need to be empirically assessed on a case by case basis, and admissibility will need to be considered on a case by case basis.

Although we have concentrated on admissibility under [FRE 702](#) and *Daubert*, and to a lesser extent *Frye*, our arguments are founded on what we consider to be good scientific practice, and, from a scientific perspective, these should be relevant irrespective of the legal standard for admissibility.

Although our focus has been on the admissibility of forensic voice comparison testimony, we believe that it would be logically consistent to apply the same criteria in considering the admissibility<sup>420</sup> of testimony based on comparison of other items of forensic interest.

## X. ACKNOWLEDGMENTS

This work was partially funded by two visiting fellowships from the Simons Foundation, one awarded to each author. The authors would like to thank the Isaac Newton Institute for Mathematical Sciences for its hospitality during the program Probability and Statistics in Forensic Science which was supported by EPSRC Grant Number EP/K032208/1. The second author also received support from National Institute of Justice Award 2014-DN-BX-K032.

# TAB 5: PCAST MATERIALS

## SUMMARY OF PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY REPORT ON FORENSIC SCIENCE

The President's Council of Advisors on Science and Technology (PCAST) is a federal advisory committee created to make recommendations to the President and the Executive Office of the President in areas of science policy. PCAST was established by President George W. Bush in 2001 and reestablished in 2009 by President Barack Obama but similar scientific advisory bodies existed under previous presidents. PCAST is supported by the Office of Science and Technology Policy (OSTP), co-chaired by the Director of OSTP, and its members are scientists appointed by the President. Unless renewed, PCAST's current charter is set to expire on February 12, 2018. It is unclear whether PCAST is currently active. PCAST has issued reports on a variety of topics such as the safety of drinking water, hearing aids, and climate change.

On September 30, 2016, PCAST released a report titled *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. The PCAST report creates new criteria by which the scientific validity of pattern matching/feature-comparison forensic disciplines should be judged and argues that only those disciplines meeting these new standards of *foundational validity* and *validity as applied* should be admissible in court. The PCAST report takes the position that unless a forensic discipline has been "scientifically validated" – in other words, unless a discipline has a known error rate – then testimony associating evidence to a source in that discipline should not be admitted in court. The PCAST report concludes that firearms (ballistics), shoeprints and tire treads, complex-source DNA, and bite-marks are not sufficiently validated (and the report implies that hair and other disciplines are also not validated).

While there is much in the PCAST report with which the Department agrees, in particular the desire for increased attention and funding for forensic science, the Department believes that the report includes a seriously flawed legal analysis on admissibility of expert forensic science testimony. This is not a new position. Attorney General Lynch issued a statement rejecting the report's recommendations on admissibility at the time the report was issue (see below). The Department also takes issue with much of the application of the science as reported in the PCAST report. The report purports to reflect scientific consensus rather than advocacy or opinion but makes broad assertions about science, forensic practice, and the law. The report has naturally led to motions to exclude any forensic testimony related to ballistics and other disciplines by defense counsel across the country. These challenges have largely been rejected by the courts that have ruled on them but resources from prosecutors and forensic laboratories have been expended to respond to them.

### ATTORNEY GENERAL LYNCH STATEMENT ON PCAST PROVIDED TO WALL STREET JOURNAL (09/20/2016)

*Over the past several years, the Department of Justice has taken unprecedented steps to strengthen forensic science, including new investments in forensic science research, draft guidance to lab experts when they testify in court, and reviews of forensic testimony in closed cases. We remain confident that, when used properly, forensic science evidence helps juries identify the guilty and clear the innocent, and the Department believes that the current legal standards regarding the admissibility of forensic evidence are based on sound science and sound legal reasoning. We understand that PCAST also considered the issue of certain legal standards, alongside its scientific review. While we appreciate their contributions to the field of scientific inquiry, the Department will not be adopting the recommendations related to the admissibility of forensic science evidence.*



UNT Center for Human Identification

June 17, 2017

To whom it may concern:

When the President's Council of Advisors on Science and Technology (PCAST) Report first was published in 2016, it was obvious that the report was not particularly helpful from a scientific perspective as it was myopic, full of error, and did not provide data to support its contentions. A more significant concern regarding the failings of the PCAST Report was that it claimed its focus was on science, but obviously was dedicated substantially to policy. Initially I considered writing a critique about the failings of the PCAST Report to assist the community. But the problems with this report were so obvious that I did not think it would be necessary to devote time to such an effort. Indeed my prediction was correct in that the report would be (and has been) rejected by the scientific community as well as overwhelmingly by the courts. However, the PCAST Report is being relied on by the Public Defender Service in U.S. v. Benito Valdez (Motion to Exclude the Testimony of the Government's proposed expert witness in Firearms Examination and Memorandum of Points and Authorities in Support, dated June 2, 2017) as a scientifically sound review of the state of the forensic sciences. Therefore, it has become necessary to address the serious limitations of the PCAST Report and convey that it is an unsound, unsubstantiated, non-peer-reviewed document that should not be relied upon for supporting or refuting the state of the forensic sciences.

My credentials to be able to opine on the failings of the PCAST Report are based on my work of more than 30 years in research, development, validation, and implementation of DNA typing methodologies for forensic applications (my CV is attached). I received a Ph.D. in Genetics in 1979 from Virginia Polytechnic Institute and State University. From 1979-1982, I was a postdoctoral fellow at the University of Alabama at Birmingham and carried out research predominately on genetic risk factors for such diseases as insulin dependent diabetes mellitus, melanoma, and acute lymphocytic leukemia. In 1983, I joined the research unit at the FBI Laboratory Division to carry out research, development, and validation of methods for forensic biological analyses. The positions I held at the FBI include: research chemist, program manager for DNA research, Chief of the Forensic Science Research Unit, and the Senior Scientist for the Laboratory Division of the FBI. I have contributed to the fundamental sciences as they apply to forensics in analytical development, population genetics, statistical interpretation of evidence, and in quality assurance. Some of my technical efforts have been: 1) development of analytical assays for typing myriad protein genetic marker systems, 2) designing electrophoretic instrumentation, 3) developing molecular biology analytical systems to include RFLP typing of VNTR loci and PCR-based SNP, VNTR and STR assays, and direct sequencing methods for mitochondrial DNA, 4) new technologies such as use of massively parallel sequencing; and 5) designing image analysis systems. I worked on laying some of the foundations for the current

statistical analyses in forensic biology and defining the parameters of relevant population groups. I have published approximately 600 articles (more than any other scientist in the area of forensic genetics), made more than 730 presentations (many of which were as an invited speaker at national and international meetings), and testified in well over 250 criminal cases in the areas of molecular biology, population genetics, statistics, quality assurance, validation, and forensic biology. In addition, I have authored or co-authored books on molecular biology techniques, electrophoresis, protein detection, forensic genetics, and microbial forensics. I was directly involved in developing the quality assurance standards for the forensic DNA field in the United States. I have been a chair and member of the Scientific Working Group on DNA Methods, Chair of the DNA Commission of the International Society of Forensic Genetics, and a member of the DNA Advisory Board. I was one of the original architects of the CODIS National DNA database, which maintains DNA profiles from convicted felons, from evidence in unsolved cases, and from missing persons.

Some of my efforts over the last 16 years also are in counter terrorism, including identification of victims from mass disasters, microbial forensics and bioterrorism. I was an advisor to New York State in the effort to identify the victims from the WTC attack. In the area of microbial forensics, I was the chair of the Scientific Working Group on Microbial Genetics and Forensics, whose mission was to set QA guidelines, develop criteria for biologic and user databases, set criteria for a National Repository, and develop forensic genomic applications. I also have served on the Steering Committee for the Colloquium on Microbial Forensics sponsored by American Society of Microbiology, was an organizer of four Microbial Forensics Meetings held at The Banbury Center in the Cold Spring Harbor Laboratory, and participated on several steering committees for NAS sponsored meetings.

In 2009 I became Executive Director of the Institute of Applied Genetics and Professor at the University of North Texas Health Science Center at Fort Worth, Texas. I currently direct the Center for Human Identification. I also direct an active research program in the areas of human forensic identification, microbial forensics, emerging infectious disease, human microbiome, molecular biology technologies, and pharmacogenetics (or molecular autopsy). I also currently am an appointed member of the Texas Forensic Science Commission.

Of note, the PCAST Committee relied on my work and as a noted expert which is supported by the report's citation of my work several times all in a favorable manner. Indeed, I am the scientist at the FBI that is mentioned as Dr. Lander's co-author to bolster his credentials in the forensic sciences (see footnotes 17 and 20). My work is cited in footnotes 33, 149, 183, 185, 187, and 209.

The report lacks scientific substance. It is cloaked with a veneer of science but in actuality is an attempt to set policy. The report discusses and advocates validation (a topic all should agree is important). Yet the topic is only addressed superficially providing definitions that already are well known with generalizations and terms it calls criteria. Nothing novel was provided by the report (see examples in references 1-7 that already have discussed the same criteria but to a greater degree than in the report). Moreover, the report does not provide any substantial guidance on how to perform validation studies for any of the disciplines it addresses. There are basic validation criteria such as sample size, power analyses, types of samples, sensitivity, specificity, dynamic range, purity of analyte, etc. that the report does not address per se or only touches upon (and instead uses black box studies for its only endeavor into sampling uncertainty and for a



misguided attempt at addressing the potential for error). The PCAST Committee could have done a service to the community if it had selected some validation studies that it claims to have reviewed (although such claims are suspect as there is no documentation supporting the claims) and described specifically those studies that the PCAST Committee deemed inappropriate and/or inadequate. Then, the PCAST Committee could have laid out how those studies should have been performed with the real substantive criteria and examples that are necessary to perform a validation study. Leading by example would have been helpful; instead the report just dismisses most of the work performed in 2000 plus articles that it claims (sic) to have reviewed. The report criticizes the forensic community for a lack of validation studies but does not describe what is lacking in any substantive way.

The Report does not describe data from each of the disciplines that could be relied upon. It is difficult to believe that in 2000 papers, the PCAST Committee claims to have relied upon, that there are no data of value. There are no indications that the PCAST Committee actually assessed the data in the literature. There is little if any documentation in this regard which should be extremely troubling to all given the PCAST Committee's strong positions of the importance of validation, documentation, and peer-reviewed publication for the forensic science community. The PCAST Committee clearly takes a "do as I say, not as I do" position. The report contains no discussion on the criteria that were used to assess the literature, the criteria that were used to dismiss the literature as inadequate, and no documentation that any data (if existing) are readily available to support that the PCAST committee performed a sound, full and complete review. Again, these issues are most disconcerting because it is apparent that the PCAST Committee in its undertaking did not hold itself up to the same standards of validation, documentation, and peer-review that it espouses the forensic community should embrace (compounded as a number of the criticisms in the report are unfounded). The report provides some guidance on basic statistics, such as estimating false positive rates (which are not novel). However, this lecturing on proper statistics is troubling to say the least as the report misuses statistics in its own cursory efforts.

The following are examples from the report to support my above claims. They are not comprehensive as it is unnecessary to go page-by-page to indicate the serious problems with the PCAST Report. A few examples should suffice to demonstrate why this report has been so underwhelming and been ignored by most scientists and the courts. In pointing out the failings of the report I will focus on topics that transcend the disciplines and specifically on my area of expertise, i.e., DNA; I could not adequately address the other disciplines and what data do or do not exist in those forensic science areas. I leave specifics of other disciplines to those with requisite expertise. However, I stress that since the report misinforms on forensic DNA applications, which is considered the "gold standard" and well-documented in the scientific literature (even the report acknowledges that), then there is a strong indication that perhaps the report missed the mark on the other disciplines as well.

I take the position that improvements in forensic sciences are needed. Indeed, all science continues to improve. It is never static. In my field of DNA typing, I and others have been and currently are working on developing better/improved methods, such as the use of next generation sequencing and new software tools. It would be improper to say that any method is perfect and cannot be made better. That position, though, is not a wholesale condemnation of the forensic sciences. Each discipline, or better yet each application, should be assessed in context as a holistic system (not solely based on validation as the report seemingly myopically espouses) and

the types/quality of samples encountered in specific cases. The report's generalization of issues avoids addressing an extremely important question – was the analysis/interpretation in this case performed correctly?

The first two examples presented below are particularly egregious and point to the dearth of substance in the report. The report states on page 2

“In the course of its study, PCAST compiled and reviewed a set of more than 2,000 papers from various sources—including bibliographies prepared by the Subcommittee on Forensic Science of the National Science and Technology Council and the relevant Working Groups organized by the National Institute of Standards and Technology (NIST); submissions in response to PCAST's request for information from the forensic-science stakeholder community; and PCAST's own literature searches.”

On page 67 of the report it is stated

“PCAST compiled a list of 2019 papers from various sources—including bibliographies prepared by the National Science and Technology Council's Subcommittee on Forensic Science, the relevant Scientific Working Groups (predecessors to the current OSAC), and the relevant OSAC committees; submissions in response to PCAST's request for information from the forensic-science stakeholder community; and our own literature searches.”

There were two citations to support the review of the 2000 or so papers that the PCAST relied upon:

[www.nist.gov/forensics/workgroups.cfm](http://www.nist.gov/forensics/workgroups.cfm).

[www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensics\\_references.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_references.pdf).

Neither of these sites appear to show (or allow for ready identification) what those articles were that the PCAST Committee reviewed and then relied upon. More so, there are no criteria and no data in the report or at these sites on what the PCAST Committee actually read, noted, reviewed, quantified, calculated, accepted, rejected, and/or debated. The report advocates emphatically and repeatedly the virtues of validation, documentation, and peer-review. Yet the report does not contain such information and thus does not meet as a minimum the requirements that it lambasted the forensic science community for lacking. This inconsistency between recommended requirements and lack of performance by the PCAST Committee is most noted as there is substantial documentation in the forensic science community (in many disciplines) but not in this report.

This lack of documentation should be considered in light of the report's statements on pages 1 and 22

“PCAST concluded that there are two important gaps: (1) the need for clarity about the scientific standards for the validity and reliability of forensic methods and (2) the need to

evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable.”

The report also states on pages 4 and 21

“It is the proper province of the scientific community to provide guidance concerning scientific standards for scientific validity, and it is on those *scientific* standards that PCAST focuses here.”

Yet the PCAST Committee did not provide its data to support the validity of its own work. There simply is no accounting of the PCAST Committee’s work to demonstrate it assessed the 2000 papers and how it came to the conclusions it rendered.

This evident failing is exacerbated by the reports statement on page 6

“The forensic examiner must have been shown to be *capable* of reliably applying the method and must *actually* have done so. Demonstrating that an expert is *capable* of reliably applying the method is crucial—especially for subjective methods, in which human judgment plays a central role. From a scientific standpoint, the ability to apply a method reliably can be demonstrated only through empirical testing that measures how often the expert reaches the correct answer. Determining whether an examiner has *actually* reliably applied the method requires that the procedures actually used in the case, the results obtained, and the laboratory notes be made available for scientific review by others.”

No one knows what method(s) the PCAST Committee used; but it is clear that it did not hold itself to the same standard either by *capability* or *actually* performing. This report cannot be held up for scientific review (as indicated on page 6 of the report – see immediately above). There are no notes or results available.

As the report says repeatedly (see pages 6 and 32)

“We note, finally, that neither experience, nor judgment, nor good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) can substitute for actual evidence of foundational validity and reliability.”

The academic and professional standings of the PCAST Committee members are not a substitute for good practices (none of which are documented). No one should take seriously this report because it has little substance to support its contentions.

The second most egregious example is the misuse and disregard for statistics. It may appear to the casual observer that the PCAST Committee is steeped in statistics and thus all statistics presented must be meaningful. For example, the report dedicates Appendix A for some discussion on statistics. But this guidance is rather basic and not particularly helpful to guide the community for any specific discipline or application. Yet when it comes to substance the PCAST Committee fails again which is evident in its own use of statistics. Consider the statements in the report on page 3

“Reviews by the National Institute of Justice and others have found that DNA testing during the course of investigations has cleared tens of thousands of suspects and that DNA-based re-examination of past cases has led so far to the exonerations of 342 defendants. Independent reviews of these cases have revealed that many relied in part on faulty expert testimony from forensic scientists who had told juries incorrectly that similar features in a pair of samples taken from a suspect and from a crime scene (hair, bullets, bitemarks, tire or shoe treads, or other items) implicated defendants in a crime with a high degree of certainty.”

Then on page 26

“DNA-based re-examination of past cases, moreover, has led so far to the exonerations of 342 defendants, including 20 who had been sentenced to death, and to the identification of 147 real perpetrators.”

A similar statement is found on page 44 (footnote 94). These findings appear to support the assertion on page 44 of the report

“It is *important* because it has become apparent, over the past decade, that faulty forensic feature comparison has led to numerous miscarriages of justice.”

I do not dispute that there have been 342 post-conviction exonerations. I am not sure what the number of exonerations is when the report says “many relied in part on faulty expert testimony” – because the report does not quantify what is meant by many. However, one wrongful analysis or testimony is one too many, and every effort should be made to minimize forensic science errors. The exoneration of 342 convicted felons is serious and topic in its own right (and again way too many). But this number is statistically meaningless and out of context. The PCAST Committee should have recognized this obvious aspect of the use of numbers. The PCAST Committee did not perform any statistical analyses or even appear to collect the data necessary to put these numbers in proper perspective. The PCAST Committee should have identified how many cases in total that have been reviewed to date (especially given that the report discusses the proper way to calculate a false positive rate, the Committee does not follow through with the same verve). This number of 342 may be and is likely a very small percentage of the total number of cases reviewed, especially since the innocence project has been around for 25 years (see <https://25years.innocenceproject.org/>). Moreover, the PCAST Committee did not convey how many post-conviction analyses that have been performed over the past 25 years in which there was no evidence of improper scientific performance, findings or faulty testimony. It would seem that such obvious basic information eluded the PCAST Committee. Those cases that were reviewed over the past 25 years in which no misuse of forensic science analyses were detected would indicate that perhaps the forensic science field is not so scientifically corrupt as the report implies. More so it would indicate that proper results can be obtained (at least most of the time).

The report discusses error rates substantially using statements such as on page 6

“Similarly, an expert’s expression of *confidence* based on personal professional experience or expressions of *consensus* among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies.”

The PCAST Report also recommends

“For subjective feature-comparison methods, because the individual steps are not objectively specified, the method must be evaluated as if it were a “black box.”

Smrz et al (8) (a paper of which I am a co-author) recommended the black box approach after the review of the FBI Laboratory’s latent print misidentification related to the Madrid bombing incident, and the PCAST Report advocates the use of such black box studies. I concur that a black box approach has some value but strongly caution that one must consider the proper utility of such studies. The authors of the PCAST Report calculated upper bound error rates based on the results of the very few black box studies they discuss; the PCAST Committee seemingly implies that these upper bound error rates are somehow meaningful to report in every case analysis. A black box study can demonstrate generally whether or not a method can yield reliable results where a human is substantially involved in the interpretation of results. But it does not necessarily help address error that may or may not have occurred during a specific case analysis.

There are several problems with such a simplistic generalization that the authors of the PCAST Report have taken regarding use of black box studies. A black box study only tests those individuals involved in the study. Therefore, the performance of the rest of the analysts of the forensic science community is not covered by the study, and the results of the study may not apply to those analysts. Some individuals perform better than others in black box studies. The average rate inflates the performance of the poorer analysts and deflates the performance of the better analysts tested in the study. Therefore, the error rate values calculated by the PCAST authors likely do not apply to most analysts. Moreover, the information content and quality of results from a forensic science analysis vary from sample to sample. Treating all sample results equally and applying a single error rate does not convey the chance for error in a particular analysis. As the PCAST Report states (see below) DNA mixture interpretation is more challenging than interpretation of single source DNA profiles. If the PCAST Committee recognizes that differences in the quality of DNA evidence affect difficulty of interpretation, then the PCAST Committee should have been able to realize that the same holds for black box study results and different quality evidence (another obvious inconsistency in the report).

A known error rate or proficiency test mistake is at best some indirect measure of the verity of the proposed results in any given sample analysis, but can never be a direct measure of the reliability of the specific result(s) in question (9). Consider a hypothetical crossing of a street where there is a 1% error (arbitrary for sake of discussion) of being hit by a car. At the beginning of the journey crossing the road there is a 1% error of being hit. While crossing the road the chance can increase or decrease depending on circumstances (possibly being greater at the center of the road and less within lanes). If the individual successfully crosses the road, then the error drops to zero. Of course, different roads (such as a busy interstate vs a rural back road) have different *a priori* chances of error (i.e., similar to the quality of evidence affects the degree of difficulty). Ultimately the issue of crossing the road is did the individual successfully cross the road or get hit. The same holds for casework, i.e., is there an error or is there not an error in the performance or analysis. Given that the black box studies mentioned in the report did have a good degree of success, there is support that a process can generate a reliable result. Thus it still comes back to determining if an error of consequence was committed in a specific case. Oddly not mentioned in the PCAST Report is that most of the forensic disciplines addressed carry out non-consumptive forms of examination. Therefore, the most direct way to measure the truth of

the purported results is to have another expert conduct his/her own review, as is advocated by the National Research Council Report II for DNA analyses (10). Re-analysis would be more meaningful instead of espousing hypothetical error rates, which may not apply to the actual results and/or analysts involved. Indeed, the above mentioned black box studies and the missing data on total number of cases from innocence project case reviews do support that tests can yield reliable results but that most of the problems (as discussed below for DNA mixtures) have been due to misapplication. Therefore, case peer-review can be an effective approach to identify misapplications. However, the PCAST Report seems to ignore the value of this practice which demonstrates the reports myopic assessment of the forensic sciences and lack of consideration of a holistic systems approach.

The PCAST Report singles out validation as essentially the sole basis for reliability. Instead under a systems approach there are several components that impact an outcome, and the reliance on these several features increases validity and reliability in any one case. Quality performance is an essential component for obtaining reliable results and for reducing the chance of error. Quality assurance provides an infrastructure to promote high performance, address errors that arise, and improve processes. In addition to validation studies, there are other mechanisms such as technical review of a case that reduce error. This technical review is performed within the laboratory before issuing a report and also outside the laboratory when an expert witness is acquired by the opposing side to assess results and interpretations. The PCAST Report seems to ignore the value of these additional quality measures and the strength of the adversary system. Error rates are difficult to calculate; they are fluid. When an error of consequence (i.e., a false “match”) occurs, under a sound quality assurance program corrective action is taken (to include review of cases analyzed by the examiner prior to and post the discovery of the error). When the corrective action is such that the individual will no longer commit that error, it no longer impacts negatively on the individual’s future performance. In fact, he/she is better educated and less likely to err. The calculation of a current error rate then should not include past error(s). Having said that, past error should not be ignored; if desired, it could be raised in court or other deliberations. The defense (or prosecution), if it believes it useful, should make use of such information during a cross-examination of an expert. But the PCAST Report does not address the shortcomings of the calculated error rate as it uses it; it treats the upper bound error rate calculation from black box studies as if they are robust and specific (which they are not).

Notably the PCAST Report tends to dismiss experience and judgment, implying it has little value. I agree that experience and judgment standing alone should be considered with caution. However, the vast majority of forensic science disciplines work in a systems approach, i.e., many facets to the process; experience is but one factor among several to effect a quality result. Even though the PCAST Report dismisses experience it again shows its inconsistencies about the province of experience. Consider the following statements on page 55 of the report

“In some settings, an expert may be scientifically capable of rendering judgments based primarily on his or her “experience” and “judgment.” Based on experience, a surgeon might be scientifically qualified to offer a judgment about whether another doctor acted appropriately in the operating theater or a psychiatrist might be scientifically qualified to offer a judgment about whether a defendant is mentally competent to assist in his or her defense.”

“By contrast, “experience” or “judgment” cannot be used to establish the scientific validity and reliability of a metrological method, such as a forensic feature-comparison method. The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of “judgment.” It is an empirical matter for which only empirical evidence is relevant. Moreover, a forensic examiner’s “experience” from extensive casework is not informative—because the “right answers” are not typically known in casework and thus examiners cannot accurately know how often they erroneously declare matches and cannot readily hone their accuracy by learning from their mistakes in the course of casework.”

Even to a lay person these statements should be obviously inconsistent, troubling and point to the inadequacy of the PCAST Committee addressing the topic of forensic science reliability. I fail to see why the medical and psychology fields can have another expert review another’s work (on what may be life and death decisions) and opine on the analyses/interpretations; yet a qualified forensic science analyst cannot perform a technical review of forensic work to assess analyses/interpretations (especially since the report has ignored data that support that at some level forensic testing is reliable). The logic of the PCAST Committee escapes me.

The PCAST Report discusses DNA typing and the limitations that have been encountered with mixture interpretation. For example on page 75 the report states

“DNA analysis of complex mixtures—defined as mixtures with more than two contributors—is inherently difficult and even more for small amounts of DNA.”

I concur that it is more challenging to interpret DNA mixtures compared with single-source DNA profiles. But the report fails to add that difficult does not necessarily translate into impossible or that proper interpretations can be made. The difficulties with mixture interpretation were not due to a lack of good, valid approaches to employ as there were valid approaches and also not due to the fact that there is some subjective judgment with interpretations. The issue, and it is a serious one, was that many of the practitioners in the forensic DNA community were inadequately trained, did not seek out solutions, or instead chose to wait for guidance (see pages 77-78 of the PCAST report and discussion on Texas and mixture interpretation). These issues were similar to the mixture interpretation problems at the Department of Forensic Sciences in Washington, DC (in which I was the scientist who identified the problems).

The PCAST Report assails the use of the Combined Probability of Inclusion (CPI) which is one of the methods used by the community and endorsed by the DNA Advisory Board (11) 17 years ago. However, the discussion of the Texas Forensic Science Commission (TFSC) (of which I was deeply involved in the review of mixture interpretation for the State) and how it pursued and addressed inappropriate interpretation of mixtures actually implies that valid methods do exist; otherwise how could a group of international experts (of which I was one of the experts) assess the situation, determine that there are problems in the application of interpretation guidelines, and provide guidance to the community to implement sound procedures?

The PCAST Committee on page 78 of the report states

“The TFSC also convened an international panel of scientific experts—from the Harvard Medical School, the University of North Texas Health Science Center, New Zealand’s



forensic research unit, and NIST—to clarify the proper use of CPI. These scientists presented observations at a public meeting, where many attorneys learned for the first time the extent to which DNA-mixture analysis involved subjective interpretation. Many of the problems with the CPI statistic arose because existing guidelines did not clearly, adequately, or correctly specify the proper use or limitations of the approach.”

The report properly focuses on lack of detailed guidelines on interpretation and does not suggest that the principles of how to calculate the CPI are erroneous. Indeed, nowhere in the report are there any data to indicate that the CPI is foundationally erroneous.

Yet, the report then states on page 78

“In summary, the interpretation of complex DNA mixtures with the CPI statistic has been an inadequately specified—and thus inappropriately subjective—method. As such, the method is clearly not foundationally valid.”

The allegation that the CPI is not foundationally valid demonstrates the lack of understanding (and again the lack of documentation of review) by the PCAST Committee. In fact, these statements also demonstrate another report inconsistency – this time about the principles of statistical calculations related to DNA profiles. On page 72 the report states

“The process for calculating the random match probability (that is, the probability of a match occurring by chance) is based on well-established principles of population genetics and statistics.”

The random match probability is one approach to calculating a statistic for single-source samples and appears to be endorsed by the PCAST Committee as well-established and thus valid. Yet, the PCAST Committee takes the opposite position for the CPI stating it is not foundationally valid. If one reads my colleagues and my most recent paper on the CPI (12), cited in the PCAST Report, it is clear that the principles of the foundational validity of the CPI are the same as those for the random match probability. Consider a similar situation which is the chance of drawing four aces in a row from a standard deck of cards is estimated to be 1 in 270,275. This value is based on probability theory and does not require an empirical testing to be published in the peer reviewed literature to support its validity. The CPI and random match probability use the same population frequency data and the same well-established principles of population genetics and statistics. While this is another example of myopia by the PCAST Committee, it borders on the bizarre that the PCAST Committee failed to understand the foundations of DNA statistics.

All know the PCAST Committee had access to the most recent paper on the use of the CPI (and the references within that paper) as it is stated on page 78 of the report

“Because the paper appeared just as this report was being finalized, PCAST has not had adequate time to assess whether the rules are also *sufficient* to define an objective and scientifically valid method for the application of CPI.”

I note that the CPI is a rather simple concept and its foundations are basic. It is surprising that the PCAST Committee, which touts its vast expertise, could not readily assess the paper. Given the importance of their report and this topic it also is surprising that they would not have done so before finalizing their report.

The PCAST Report recognizes that probabilistic genotyping is an advancement to improve or reduce subjectivity in DNA mixtures (see page 79). I concur. But the report states on page 79

“Appropriate evaluation of the proposed methods should consist of studies by multiple groups, *not associated with the software developers*, that investigate the performance and define the limitations of programs by testing them on a wide range of mixtures with different properties.”

Also the report states on page 81

“Because empirical evidence is essential for establishing the foundational validity of a method, PCAST urges forensic scientists to submit and leading scientific journals to publish high-quality validation studies that properly establish the range of reliability of methods for the analysis of complex DNA mixtures.”

Publication is part of the peer-review process and I support publication by the developers and others who adopt the method. But the PCAST Committee has placed a requirement that is unrealistic to meet which is publication by the user laboratories. It is likely that a few at most laboratories will be able to publish their validation testing of the software. Anyone who serves on editorial boards of scientific journals should know that journals are unlikely to publish additional studies because they are not considered novel. Yet, the PCAST Committee failed to recognize this fact.

It is important to stress that the report contains no criticisms of probabilistic genotyping and still there are no data contained in the report that demonstrate that the PCAST Committee actually reviewed (or better yet tested) the current probabilistic genotyping software programs (even though it claims to have done extensive review, such as the undocumented 2000 papers).

Forensic laboratories are required to perform validation studies, and there are substantial data on mixtures that support the validity of mixture interpretation and use of probabilistic genotyping. Mixture studies are required to be performed by every laboratory engaged in analyzing such evidence as part of their validation studies. Many of these studies lack novelty and thus will never be published in peer-review journals. However, the PCAST Committee could have contacted a number of forensic DNA laboratories who have implemented one of the probabilistic genotyping software programs (as there were laboratories operating or near implementation of the tools at the time of the report’s publication) to gain access to the validation data to determine whether there are sufficient data to support the already peer-reviewed published work. There is no indication that the PCAST Committee made any effort to become informed to opine on the reliability and validity of probabilistic genotyping.

The PCAST Committee simply ignored a wealth of validation data residing in crime laboratories. If the PCAST Committee had taken a holistic approach, they would have considered the totality of data in determining whether there is support for the validity and reliability of probabilistic genotyping. Peer-review publications by the developers and validation data by the users combined clearly support the software and its applications. Indeed, this failure of the PCAST Committee of not considering all available data is reminiscent of a similar situation that occurred 25 years ago with another report – the National Research Council I Report (NRC I) (13). The NRCI Report proposed a non-scientific, *ad hoc* way to calculate statistics called the ceiling principle. The ceiling principle had no genetics foundation or validity and was roundly rejected. One of the bases for the proposed ceiling principle approach (espoused by the NRC I Committee) was a lack of population data. There were substantial population data in crime

laboratories world-wide at the time the NRC I Report was published; but the NRC I Committee did not seek out the data. As soon as the NRC I Report was published, I reached out to my colleagues around the world and gathered the existing data which were then compiled into a five volume compendium (14). If the NRC I Committee had chosen to consider extant population data, they might have prepared a more informed Report. The outcome was that the National Academy of Sciences convened a second committee and produced the sound NRC II Report (10), which was steeped in fundamental population genetics and statistical applications. The findings of the NRC II Report in part were based on the data I compiled in the five volume compendium which were available prior to the publication of the rejected NRC I Report. The PCAST Report has taken the same blinded approach and ignored extant data with a similar outcome as 25 years ago – a report that provides little value for assessing the state-of-the-art and even less value for providing guidance to improve the forensic sciences.

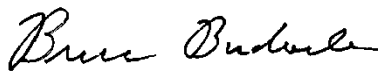
In conclusion, the few examples above demonstrate that the PCAST Report 1) is **not** scientifically sound, 2) is **not** based on data, 3) is **not** well-documented, 4) misapplies statistics, 5) is full of inconsistencies, and 6) does **not** provide helpful guidance to obtain valid results in forensic analyses.

## References

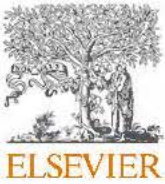
1. Budowle, B. and SWGMGF Members: Quality assurance guidelines for laboratories performing microbial forensic work. *Forens. Science Communications* October 5(4): 2003, At: [www.fbi.gov/hq/lab/fsc/current/2003\\_10\\_guide01.htm](http://www.fbi.gov/hq/lab/fsc/current/2003_10_guide01.htm).
2. Budowle, B., Schutzer, S.E., Einseln, A., Kelley, L.C., Walsh, A.C., Smith, J.A.L., Marrone, B.L., Robertson, J., and Campos, J.: Building microbial forensics as a response to Bioterrorism. *Science* 301: 1852-1853, 2003.
3. Budowle, B., Schutzer, S.E., Morse, S.A., Martinez, K.F., Chakraborty, R., Marrone, B.L., Murch, R.S., Jackson, P.J., Williamson, P., Harmon, R., and Velsko, S.P.: Guidance for validation in microbial forensics. *Applied and Environmental Microbiology* 74:5599-5607, 2008.
4. Budowle, B., Connell, N.D., Bielecka-Oder, A., Colwell, R.R., Corbett, C.R., Fletcher, J., Forsman, M., Kadavy, D.R., Markotic, A., Morse, S.A., Murch, R.S., Sajantila, A., Schmedes, S.E., Ternus, K.L., Turner, S.D., Minot, S.: Validation of high throughput sequencing and microbial forensics applications. *BMC Invest. Genet.* 5:9, 2014.
5. The FBI Director's Forensic Quality Assurance Standards for DNA Testing Laboratories, [http://media.wix.com/ugd/4344b0\\_4a22824ce56f43d4b1a4d2486409f95d.pdf](http://media.wix.com/ugd/4344b0_4a22824ce56f43d4b1a4d2486409f95d.pdf)
6. SWGDAM Validation Guidelines for Forensic DNA Analysis Methods, [https://docs.wixstatic.com/ugd/4344b0\\_813b241e8944497e99b9c45b163b76bd.pdf](https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf)
7. SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems, [https://docs.wixstatic.com/ugd/4344b0\\_22776006b67c4a32a5ffc04fe3b56515.pdf](https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf)

8. Smrz, M.A., Burmeister, S.G., Einseln, A., Fisher, C.L., Fram, R., Stacey, R.B., Theisen, C.E., and Budowle, B.: Review of FBI latent print unit processes and recommendations to improve practices and quality. *J. Forens. Ident.* 56(3):402-434, 2006.
9. Harmon R. and Budowle, B.: Questions about forensic science. *Science* 311:607, 2006.
10. NRC II Report, *The Evaluation of Forensic DNA Evidence*, 1996, <https://www.nap.edu/catalog/5141/the-evaluation-of-forensic-dna-evidence>
11. *Statistical and Population Genetics Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated From Pertinent Population Database(s)*, DNA Advisory Board, *Forensic Science Communications* 2(3), 2000, <https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm>
12. Bieber, F.R., Buckleton, J., Budowle, B., Butler, J., and Coble, M.D.: Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genetics* 17(1):125, 2016.
13. NRC I Report, *The Evaluation of Forensic DNA Evidence*, 1992, <https://www.nap.edu/read/5141/chapter/1#ii>
14. *Federal Bureau of Investigation: VNTR Population Data: A Worldwide Study, Volumes I-IV*, Forensic Science Research and Training Center, FBI Academy, Quantico, Virginia, 1993.

I declare under penalty of perjury that the forgoing is true and correct to the best of my knowledge.



Bruce Budowle, Ph.D.  
Director  
Center for Human Identification  
University of North Texas Health Science Center  
Fort Worth, Texas 76107  
Email: (b) (6)  
Tel: (b) (6)



Contents lists available at ScienceDirect

# Forensic Science International

journal homepage: [www.elsevier.com/locate/forsciint](http://www.elsevier.com/locate/forsciint)



## Review Article

### Finding the way forward for forensic science in the US—A commentary on the PCAST report



I.W. Evett<sup>a,\*</sup>, C.E.H. Berger<sup>b</sup>, J.S. Buckleton<sup>c,d</sup>, C. Champod<sup>e</sup>, G. Jackson<sup>f</sup>

<sup>a</sup> Principal Forensic Services Ltd., 34 Southborough Road, Bickley, Bromley, Kent, BR1 2EB, United Kingdom

<sup>b</sup> Institute for Criminal Law and Criminology, Faculty of Law, Leiden University, PO Box 9520, 2300 RA Leiden, The Netherlands

<sup>c</sup> Environmental Science & Research Ltd, Private Bag 92021, Auckland 1142, New Zealand

<sup>d</sup> Department of Statistical Genetics, University of Washington, Box 357232 Seattle, WA 98195-7232, United States

<sup>e</sup> Ecole des Sciences Criminelles, Faculty of Law, Criminal Justice and Public Administration, Université de Lausanne, Batochime – quartier Sorge, CH-1015 Lausanne-Dorigny, Switzerland

<sup>f</sup> Abertay University, Dundee, DD1 1HG, United Kingdom

#### ARTICLE INFO

##### Article history:

Received 16 March 2017

Received in revised form 30 April 2017

Accepted 18 June 2017

Available online 26 June 2017

##### Keywords:

Forensic inference

Evidence

Comparison methods

Probability

Likelihood ratio

#### ABSTRACT

A recent report by the US President's Council of Advisors on Science and Technology (PCAST), (2016) has made a number of recommendations for the future development of forensic science. Whereas we all agree that there is much need for change, we find that the PCAST report recommendations are founded on serious misunderstandings. We explain the traditional forensic paradigms of *match* and *identification* and the more recent foundation of the logical approach to evidence evaluation. This forms the groundwork for exposing many sources of confusion in the PCAST report. We explain how the notion of treating the scientist as a black box and the assignment of evidential weight through error rates is overly restrictive and misconceived. Our own view sees inferential logic, the development of calibrated knowledge and understanding of scientists as the core of the advance of the profession.

© 2017 Elsevier B.V. All rights reserved.

#### Contents

In Memoriam .....	17
1. Introduction .....	17
2. The logical approach .....	17
2.1. Framework of circumstances .....	17
2.2. Propositions .....	17
2.3. Probability of the observations .....	17
3. The match paradigm .....	18
4. The identification paradigm .....	18
5. Misconceptions, fallacies and confusions in the PCAST report .....	18
5.1. Confusion between the match and identification paradigms .....	18
5.2. Judgement .....	19
5.3. Subjective versus Objective .....	19
5.4. Transposed conditional .....	19
5.5. "Probable match" .....	20
5.6. Foundational validity and accuracy .....	20
5.7. The PCAST paradigm .....	21
5.8. The scientist as a "black box" .....	21
5.9. Black box studies .....	21
5.10. Governance .....	22

\* Corresponding author.

E-mail address: [ianevett@btinternet.com](mailto:ianevett@btinternet.com) (I.W. Evett).



6.	Our view of the future .....	22
6.1.	Logical inference .....	22
6.2.	Calibration .....	22
6.3.	Knowledge and data .....	22
7.	Conclusion .....	23
	References .....	23

## In Memoriam

This paper is dedicated to the memory of Bryan Found who did so much to advance the profession of forensic scientist through his work on calibrating and enhancing the performance of experts under controlled conditions. He will be sorely missed.

## 1. Introduction

This paper is written in response to a recent report on forensic science of the US President's Council of Advisors on Science and Technology (PCAST) [1]. There have already been several responses to the report from the forensic community [2–7] which have resulted in an addendum to the report [8]. Our main concern is that the report (and its addendum) fails to recognise the advances in the logic of forensic inference that have taken place over the last 50 years or so. This is a serious omission which has led PCAST to a narrowly focussed and unhelpful view of the future of forensic science.

The structure of our paper is as follows. In Section 2 we briefly outline our view of the requirements imposed by logic on the assessment of the probative value of evidence. This allows us to set up a framework against which we can contrast some of the suggestions of the report. In Sections 3 and 4 we briefly explain the notions of “match” and “identification” paradigms that have underpinned much of forensic inference over the last century or so. Section 5 will point out misconceptions, fallacies, sources of confusion and improper terminology in the PCAST report. Our contrasting view of the future path for forensic science follows in Section 6.

## 2. The logical approach

Much has been written over the past 40 years on inference in forensic science. The frequency of appearance of articles, papers and books on the topic has increased markedly in recent years. Practically all of this material is founded on a logical, probabilistic approach to the assessment of the probative value of scientific observations [9,10]. The PCAST report mentions this body of work only briefly and pays scant attention to its principles [11], which we list and explain briefly as follows.

### 2.1. Framework of circumstances

*It is necessary to consider the evidence within a framework of circumstances.*

A simple example will illustrate this. Imagine that a sample<sup>1</sup> has been obtained from a crime scene which yielded a DNA profile from which the genotype of the originator of the sample has been inferred. A suspect for the crime is known to have the same genotype. Because the alleles revealed by a DNA profile will be found in different proportions in different ethnic groups, it is relevant to the assessment of the probative value of this

<sup>1</sup> The term “sample” is used generically to describe what is available for forensic examination. The term is not used here to suggest any statistical sampling process.

correspondence of genotypes that a credible eyewitness of the crime said that the offender was of a particular ethnic appearance.

It follows that, when presenting an evaluation, the scientist should clearly state the framework of circumstances that are relevant to their assessment of the probative value of the observations, with a caveat that, if details of the circumstances change, the evaluation must be revisited.

### 2.2. Propositions

*The probative value of the observations cannot be assessed unless two propositions are addressed.*

In a criminal trial, these will represent what the scientist believes the prosecution may allege and a sensible alternative that represents the defence position.<sup>2</sup> In taking account of both sides of the argument, the scientist is able to assess the evidence in a balanced, justifiable way and display to the court an unbiased approach, irrespective of which side calls the witness.

Propositions may be formed at any of at least four levels in a hierarchy of propositions [12–14]. These levels are termed offence, activity, source and sub source. We do not discuss these in any depth here. Most of the PCAST report appears to address questions at the source or sub source level. Examples of these would be:

1. Sub source: The DNA came from the person of interest (POI),<sup>3</sup> or
2. Source: This fingerprint was made by the POI.

### 2.3. Probability of the observations

*It is necessary for the scientist to consider the probability<sup>4</sup> of the observations given the truth of each of the two propositions in turn.*

The ratio of these two probabilities is widely known as the *likelihood ratio* (LR) and this is a measure of the weight of evidence that the observations provide in addressing the issue of which of the propositions is true. A likelihood ratio greater than one provides support for the truth of the prosecution proposition. A likelihood ratio less than one provides support for the truth of the defence proposition.

It cannot be sufficiently emphasized that it is the scientist's role to provide expert opinion on the probability of the *observations* given the proposition. The role of assigning a value to the probability of the *proposition* given the observations is that of the jury in a criminal trial. This probability will take account, not just of the scientific observations, but also of all of the other evidence presented at court.

<sup>2</sup> We recognise that the scientist, particularly at an early stage of proceedings, may not know the position that defence will take. It is common practice for the scientist to adopt what appears to be a reasonable proposition, given what is known of the circumstances—making it clear that this is provisional and subject to change at any time.

<sup>3</sup> A source level DNA proposition would specify the nature of the recovered material, e.g. “the semen came from the POI”.

<sup>4</sup> This could be a probability density, depending on the nature of the observations. But the principle remains unchanged.

### 3. The match paradigm

In most forensic comparisons, one of the items will be from a known origin (such as: a reference sample for DNA profiling from a particular individual; a pair of shoes from a suspect; a set of control fragments of glass from a broken window). The other will be from an unknown, or disputed origin (such as: DNA recovered from a crime scene; a footwear mark from the point of entry at a burglary; or a few small fragments of glass recovered from the clothing of a suspect). It is convenient to refer to these as the *reference* and *questioned* samples, respectively. The matter of interest to the court relates to the origin of the questioned sample. This question will be addressed scientifically by carrying out observations on both samples. These observations may be purely qualitative: such as, for example, the shapes of the loops of letters such as “y” and “g” in a passage of handwriting. They may be quantitative and discrete, such as the alleles in a DNA STR profile. Or they may be quantitative and continuous, such as the refractive index of glass fragments. The match paradigm calls for a judgement, by the scientist, as to whether or not the two sets of observations agree within the range of what would be expected if the questioned sample had come from the same origin as the reference sample. The basis for that judgement may, in the case of quantitative observations, be based on a set of pre determined criteria; but where the observations are qualitative such criteria may be vague or purely judgemental.

If the two sets of observations are considered to be outside the range of what may have been expected if the two samples had come from the same source then the result may be reported as a “non match”. Depending on the nature of the observations, this provides the basis for a strong implication that the questioned and reference samples came from different sources. In many instances this conclusion will be non controversial in the sense that prosecution and defence will be content to accept it.

However, when the result of the comparison is a “match” it does not logically follow that the two samples do share the same source or even that they are likely to be from the same source. It is possible that the two samples came from two different sources that, by coincidence, have similar properties. Throughout the history of forensic science there has been the notion – often imperfectly expressed – that the smaller the probability of such a coincidence, the greater the evidential value to be associated with the observed match. In DNA profiling, for example, we encounter the notion of a “match probability”. The implication of this approach is that the jury should assign an evidential weight that is related to the inverse of the match probability.

The logical approach has done much to clarify the rather woolly inference that historically has been associated with the match paradigm but it has also demonstrated the considerable advantages of the single stage approach implied by the assignment of weight through the calculation of the likelihood ratio, over the rather clumsy and inefficient two stage approach implied by the match paradigm. This has already been pointed out by Morrison et al. [4].

### 4. The identification paradigm

Historically, fingerprint comparison was seen to be the gold standard by which the power of any other forensic technique could be judged. The paradigm here was the notion of “identification”<sup>5</sup> or

<sup>5</sup> Kirk [15] defined the term identification as only placing an object in a restricted class. The criminalist would, for example, identify a particular mark as a fingerprint. Individualization was defined by Kirk as establishing which finger left the mark. An opinion of the kind “this latent mark was made by the finger which made this reference print” is an individualization.

“individualization” (the terms are used synonymously here). Provided that sufficient corresponding detail was observed, the outcome of a comparison between a fingerprint of questioned origin and a print taken from a known person would be reported as a categorical opinion: the two were definitely made by the same person.

So, the match and identification paradigms are related with the difference that in the latter the scientist is allowed to state that the match probability is so infinitesimally small that it is reasonable to conclude that the two items came from the same source. Historically, many examiners would have claimed that the source was established with certainty to the exclusion of all others.

The identification paradigm went largely unchallenged for many years until later in the 20th century when its logical basis was questioned (see, for example, [16] or more recently [17,18]) and also when, in a number of high profile cases, misidentifications with serious consequences were exposed.

An example of the paradigm is given in box 6, p. 137 of the PCAST report (DOJ proposed uniform language) (emphasis added).

The examiner may state that it is his/her opinion that the shoe/tire *is the source of the impression* because there is sufficient quality and quantity of corresponding features such that the examiner would not expect to find that same combination of features repeated in another source. This is the highest degree of association between a questioned impression and a known source.

The PCAST report rightly indicates that the conclusions conveying “100 percent certainty” or “zero or negligible error rates” are not scientifically defensible. Such conclusions tend to overestimate the weight to be assigned to the forensic observations.

### 5. Misconceptions, fallacies and confusions in the PCAST report

The most serious weakness in the PCAST report is their flawed paradigm for forensic evaluation. Unfortunately, the report contains more misconceptions, fallacies, confusions and improper wording. In this section we will discuss the main problems with the report.

#### 5.1. Confusion between the match and identification paradigms

This is the first source of confusion in the report. For example, from p. 90 of the report (emphasis added):

An FBI examiner concluded with “100 percent certainty” that the fingerprint *matched* Brandon Mayfield . . . even though Spanish authorities were unable to confirm the *identification*.

On p. 48 we find (emphasis added):

To meet the scientific criteria of foundational validity, two key elements are required:

(1) a reproducible and consistent procedure for (a) identifying features within evidence samples; (b) comparing the features in two samples; and (c) determining based on the similarity between the features in two samples, whether the samples should be declared to be a proposed *identification* (“*matching rule*”).

We have seen that declaring a match and declaring an identification are not the same thing. Declaring a match implies nothing about evidential weight whereas declaring an identification implies evidential weight amounting to complete certainty.

The PCAST report proposes an approach that is fusion of the match and identification paradigms. See, from p. 45/46:



Because the term “match” is likely to imply an inappropriately high probative value, a more neutral term should be used for an examiner’s belief that two samples came from the same source. We suggest the term “proposed identification” to appropriately convey the examiner’s conclusion, along with the possibility that it might be wrong. We will use this term throughout the report.

If a scientist says that the questioned and reference samples match, the immediate inference to be drawn from this (as we have explained) is that they might have come from the same source but it is also true that they might not have come from the same source. These two statements make no implication with regard to evidential weight. Weight only comes from the second stage of the paradigm which entails coming up with some impression of rarity. The identification paradigm, on the other hand, is different in that it implies a statement of certainty: the two samples certainly came from the same source.

The PCAST paradigm requires that the scientist should make a categorical statement (an identification) that cannot be justified on logical grounds as we have already explained. Most scientists would be comfortable with the notion of observing that two samples *matched* but would, rightly, refuse to take the logically unsupportable step of inferring that this observation amounts to an *identification*.

### 5.2. Judgement

The report emphasises the value of empirical data (emphasis added):

The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, *is not a matter of ‘judgment’*. It is an empirical matter *for which only empirical evidence is relevant.* ([1], p. 6)

This denial of the importance of judgement betrays a poor understanding of the nature of forensic science. We offer a simple example.

Mr POI is the suspect for a crime who was arrested at time  $T$  in location  $Z$ . Some questioned material has been found on the clothing of Mr POI which is to be compared with reference material taken from the crime scene. Denote the observations on the two samples by  $y$  and  $x$  respectively. Whichever paradigm we follow, we are interested in the probability of finding material with observations  $y$  on the clothing of Mr POI if he had nothing to do with the crime. Ideally, of course, we would like a survey carried out near to time  $T$  and in the general region of  $Z$  and of people of a socio economic group  $Q$  that would include Mr POI. But this is, of course unrealistic. What we do have is a survey of materials on clothing carried out at some earlier time  $T'$  and at another location  $Z'$  and of a slightly different socio economic group  $Q'$ . Who is to make a judgement on the relevance of this survey data to the case at hand? We would argue that this is where the knowledge and understanding of the forensic scientist is of crucial importance.

The reality is, of course, that the perfect database never exists. The council is wrong: it is most certainly *not* the case that “only empirical evidence” is relevant. Without downplaying the importance of data collections, they can only inform judgement it is judgement that is paramount and informed judgement is founded in reliable knowledge.

### 5.3. Subjective versus Objective

PCAST give their definition of the distinction between “objectivity” and “subjectivity” p. 5 footnote 3.

Feature comparison methods may be classified as either objective or subjective. By objective feature comparison methods, we mean methods consisting of procedures that are each defined with enough standardized and quantifiable detail that they can be performed by either an automated system or human examiners exercising little or no judgment. By subjective methods, we mean methods including key procedures that involve significant human judgment . . .

What is suggested is that many of the decisions be moved from the examiner to the procedure and/or software. The procedure or software will have been written by one or more people and the decisions about what models are used or how decisions are made are now enshrined in paper or code. Hence all the subjective judgements are now made by this person or group of people via the paper or code. Whereas this approach could be viewed as repeatable and reproducible, the objectivity is illusory.

In the US environment, subjectivity has been associated with bias and sloppy thinking, and objectivity with an absence of bias and rigorous thinking. It is worthwhile examining whence the fear of subjectivity arises. There is considerable proof that humans are susceptible to quite a number of cognitive effects many of which can affect judgement. We suspect that the fear is that these effects bias the decisions in ways that are detrimental to justice. Hence, it is bias arising from cognitive effects that is the enemy, not subjectivity.

If we return to the concept of enforced precision, we could assume that trials could be conducted on such a system and that the outputs could be calibrated. Such a system could be of low susceptibility to bias arising from cognitive effects. We suspect that these are the goals sought by PCAST. We certainly could support calibrating subjective judgements but we see little value in pretending that writing them down or coding them makes them objective.

### 5.4. Transposed conditional

We are concerned by the report’s poor use of the notion of probability. In particular we note in the report many instances where the fallacy of the transposed conditional either occurs explicitly or is implied. We have seen that the logic of forensic inference directs us to assign a value to the probability of the observations given the truth of a proposition. The probability of the truth of a proposition is for the jury *not* the scientist. Confusion between these two different probabilities has been called the “prosecutor’s fallacy” [19]. We prefer the term *transposed conditional* because, in our experience, the fallacy is regularly committed by prosecutors, defence attorneys, the judiciary and the media alike.

The fallacy is widespread, even though it can be grounds for a retrial if given in testimony by an expert witness. The document [20] that attempts to explain DNA statistics to defence attorneys in the US describes incorrectly a likelihood ratio for a mixture profile as:

4.73 quadrillion times more likely<sup>6</sup> to have originated from [suspect] and [victim/complainant] than from an unknown individual in the U.S. Caucasian population and [victim/complainant].” ([20], p. 52)

<sup>6</sup> We are fully aware of the distinction made in statistical theory between “likelihood” and “probability”. We believe that attempting to explain that distinction in this paper would cause more confusion than the worth of it. It is our experience that in courts of law the two terms are taken to be synonymous.

This is a classic example of the transposed conditional. It is a transposition of the likelihood ratio, which would be more correctly presented as follows:

The DNA profile is 4.73 quadrillion times more likely to be obtained if the DNA had originated from the suspect and the victim/complainant rather than if it had originated from an unknown individual in the U.S. Caucasian population and the victim/complainant.

The contrast between these two statements, though apparently subtle, is profound. The first is an expression of the probability (or odds) that a particular proposition is true – this, we have seen, is the probability that the jury must address, not the scientist.<sup>7</sup> The second considers the probability of the *observations*, given the truth of one proposition then the other, which is the appropriate domain for the expertise of the scientist. It is important to realise that the first statement is not a simple rephrasing of the second statement. Whereas the second may be a valid representation of the scientist's evaluation in a given case, the first most definitely cannot be.

Consider the following quote from the first paragraph on footwear methodology in the PCAST report ([1], p. 114):

Footwear analysis is a process that typically involves comparing a known object, such as a shoe, to a complete or partial impression found at a crime scene, to assess whether the object is likely to be the source of the impression.

This is wrong. We state again that it is not for the scientist to present a probability for the truth of the proposition that the object was the source of the impression. The scientist addresses the probability of the outcome of the comparison *if* the object were the source of the impression: this probability forms the numerator of the likelihood ratio. Just as important, of course, is the probability of the outcome of the comparison *if* some other object were the source of the impression. The latter forms the denominator of the likelihood ratio. It is the two probabilities, taken together, that determine the evidential weight in relation to the two propositions of interest to the court.

The PCAST report sentence clearly states that the objective of the footwear analysis is to present a probability for the proposition given the observations, and not for the observations given the proposition. This is clearly a transposition of the conditional.

Similarly, the scientist is not in a position to consider the probability addressed in the following ([1], p. 65 and repeated on p. 146):

... determining, based on the similarity between the features in two sets of features, whether the samples should be declared to be likely to come from the same source ...

We have seen that it is not for the scientist to consider the probability that the samples came from the same source given the observation of a “match”. It is another example of the fallacy of the transposed conditional.

This confusion is systematic in the original report and we note that it continues into the addendum ([8], p. 1) (emphasis added):

These methods seek to determine whether a questioned sample *is likely to come* from a known source based on shared features in certain types of evidence.

We have seen that this is most certainly *not* what a feature comparison should aspire to. It is not the role of the forensic

scientist to offer a probability for the proposition that a questioned sample came from a given source since this would require the scientist to take account of all of the non scientific information which properly lies within the domain of the jury.

The need for precision of language when presenting probabilities is exemplified by two quotations from the report. First, from p. 8 when talking about the interpretation of a DNA profile:

Could a suspect's DNA profile be present within the mixture profile? And, what is the probability that such an observation might occur by chance?

As we read it, this second sentence can be taken to mean:

What is the probability that such an observation would be made if the suspect's DNA were not present in the mixture?

Within the logical paradigm, this is a legitimate question to ask – it is the probability of the observations given that one of the propositions were true.

However, later in the report we find (p. 52):

the random match probability – that is, the probability that the match occurred by chance”.

There is an economy of phrasing here that obscures meaning and the reader could be forgiven for believing that the question implied by the second phrase is:

What is the probability that the two samples had come from different sources and matched by chance?

This is a probability of a proposition (the two samples came from different sources) given the observation (a match) and would imply a transposed conditional. We are aware that the council may respond that this is not at all what they meant – to which we would respond that the council should have been far more careful in its phraseology.

### 5.5. “Probable match”

In giving their definition of the distinction between “objectivity” and “subjectivity” p. 5 – see footnote 3 the report states:

*how to determine whether the features are sufficiently similar to be called a probable match.*

The council do not say what they mean by a “probable match” but it seems to us that it is another example of confusion between the match and identification paradigms. Following the match paradigm there is no such thing as a probable match – the two samples either match or they do not.

### 5.6. Foundational validity and accuracy

The report distinguishes two types of scientific validity: “foundational validity” and “validity as applied”. We confine ourselves to the first of these (p. 4):

*Foundational validity* for a forensic science method requires that it be shown based on empirical studies to be *repeatable, reproducible, and accurate*, at levels that have been measured and are appropriate to the intended application. Foundational validity, then, means that a method can, *in principle*, be reliable.

Repeatability refers to the ability of the same operator with the same equipment to obtain the same (or closely similar) results when repeating analysis of the same material. Reproducibility refers to the ability of the equipment to obtain the same (or closely similar) results with different operators. As such, both are

<sup>7</sup> In Bayesian terms, the first statement is one of posterior odds. This can be derived from the second statement either by assigning prior odds of one (which would be highly prejudicial in most criminal trials) or by making the mistake of transposing the conditional. Neither is acceptable behaviour for a scientist.

expressions of precision, which is how close each measurement or result is to the others.

Accuracy is a measure of how close one or a set of measurements is to the true answer. This has an obvious meaning when we know or could know the true answer. We could imagine some measurement such as the weight of an object where that object has been weighed by some very advanced technique and we can accept that as the “true” weight. We wish then to consider the accuracy of some other, perhaps cheaper, technique. We could assess the accuracy of this second technique by using it to weigh the object multiple times and observing the deviation of the results from the “true” weight of the object.

For some questions in forensic science, such as “How much heroin is in this seized sample?” or “How much ethanol is in this blood sample?”, the notion of the accuracy of an applied analytical technique is relevant because it is possible to assess a technique’s accuracy using trials with known quantities of heroin or ethanol. However, when it comes to answering a question such as “What is the probability that there would have been a match with a suspect’s shoe if it did not make the mark at the scene of crime?”, then there is no sense in which there is a “true answer”. The values that experts assign for such probabilities will vary depending on the specific knowledge of the experts and the nature of any databases that experts may use to inform their probabilities.

We could use a weather forecaster as an illustration. If she says that there is a 0.8 probability of a sunny day tomorrow, there can be no sense in which this is a “true” statement. Equally, if tomorrow brings rain, she is not “wrong” in any sense. Nor is she “inaccurate”. A probabilistic statement of this nature may be unhelpful or misleading, in the sense that it may lead us to make a poor decision, but it cannot be either true or false.

Once we abandon the idea of a true answer for probabilities, we are left with the difficult question of what we mean by accuracy. We suggest that the report does a disservice to the important task of calibrating probabilities by a simplistic allusion to accuracy.

The PCAST report says (p. 46):

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

We have seen that the report is wrong here—it is not a matter of “accuracy” but of evidential weight.

### 5.7. The PCAST paradigm

The PCAST report proposes an approach that is fusion of the match and identification paradigms. See, from p. 45/46:

Because the term “match” is likely to imply an inappropriately high probative value, a more neutral term should be used for an examiner’s belief that two samples came from the same source. We suggest the term “proposed identification” to appropriately convey the examiner’s conclusion, along with the possibility that it might be wrong. We will use this term throughout the report.

First, we have seen that the term “match”, if used properly, makes no implication of probative value: it implies that the two samples might have come from the same source but also might have come from different sources. This is evidentially neutral. Second, we have seen that there is no place for the “examiner’s

belief that two samples came from the same source”: it is not for the scientist to assign a probability to the proposition that the two samples came from the same source.

Next we must consider what the council understand the phrase “proposed identification” to mean. Do they mean that, because it is an identification, it is a categorical opinion? Note that the qualifier “proposed” does not make the identification less than categorical—if it were probabilistic it could not be “wrong”.<sup>8</sup> If it is not probabilistic then the scientist is to provide a categorical opinion while telling the court that he/she might be wrong! It is difficult to believe that any professional forensic scientist would be happy to be put in this position.

### 5.8. The scientist as a “black box”

On page 49 we find:

For subjective methods, procedures must still be carefully defined but they involve substantial human judgment. For example, different examiners may recognize or focus on different features, may attach different importance to the same features, and may have different criteria for declaring proposed identifications. Because the procedures for feature identification, the matching rule, and frequency determinations about features are not objectively specified, the overall procedure must be treated as a kind of “black box” inside the examiner’s head.

The report justifiably emphasises weaknesses of qualitative opinions. The intuitive “black box” view of the scientist will certainly have been true in many instances in the past and, indeed, in certain quarters in the present day. But for us the solution is emphatically not to continue to treat this as an acceptable state of affairs for the future. The PCAST view appears to be “it’s a black box, so let’s treat it like a black box”. Our approach has been, and will continue, to break down intuitive mental barriers by expanding transparency, knowledge and understanding. We do not see the future forensic scientist as an *ipse dixit* machine whatever the opinion, we expect the scientist to be able to explain it in whatever detail is necessary for the jury to comprehend the mental processes that led to it.

### 5.9. Black box studies

That the council intend the proposed identification to be categorical is clarified in the following from page 49 (emphasis added):

In black box studies, many examiners are presented with many independent comparison problems—typically, involving “questioned” samples and one or more “known” samples—and asked to declare whether the questioned samples came from the same source as one of the known samples.<sup>9</sup> The researchers then determine how often examiners reach erroneous conclusions.

PCAST proposes that the error rates from such experiments would be used to assign evidential value at court.

We are strongly against the notion that the scientist should be forced into the position of giving categorical opinions in this way. Whereas, we are strongly in favour of the notion of calibrating the

<sup>8</sup> Though, of course, it would be logically incorrect because it would imply a transposed conditional.

<sup>9</sup> In footnote 111 the report says: “Answers may be expressed in such terms as “match/no match/inconclusive” or “identification/exclusion/inconclusive”. This strengthens our belief that the council see match and identification as interchangeable”.

opinions of forensic scientists under controlled conditions we see those opinions expressed in terms of statements of evidential weight. We return to the subject of calibration later.

### 5.10. Governance

PCAST suggests that forensic science should be governed by those, such as metrologists, from outside the profession. This speaks to the view, reinforced by a very selective reference list, that the forensic science discipline is not to be trusted with developing procedures, testing them, and self governance. We do not reject input from outside the profession: we welcome it. But our own observations are that those outside may be engaged to different extents, varying from a passing interest to years of study. They may be unduly influenced by headlines in newspapers highlighting or exaggerating deficiencies. On occasion, these same commentators from outside the profession may not recognise the limitations in their own knowledge base where it concerns specifically forensic aspects, may be reticent to consult subject matter experts from amongst practising scientists and may give well intentioned, but erroneous, advice [1,21].

## 6. Our view of the future

### 6.1. Logical inference

The recommendations of the PCAST report are founded on a conflation of two classical forensic paradigms: match and identification. These paradigms are as old as forensic science but their inadequacies and illogicalities have been comprehensively exposed over the last 50 years or so. All of us maintain, and have done so in our writings, that the future of forensic science should be founded first on the notion of logical inference and second on the notion of calibrated knowledge. The former leads to a framework of principles (which have been adopted by ENFSI) and we are disappointed that PCAST has apparently chosen to ignore, or at most pay lip service to, this fundamental change. The second is a deeper and far richer concept than the profoundly limited notion of false positive and false negative error rates: this is the notion of *calibration*.

### 6.2. Calibration

We are most definitely in favour of the studying of expert opinion under controlled circumstances, see for example Evett [22] but proficiency testing is far more than the counting of errors. The PCAST black box approach calls for a categorical opinion that is recorded as right or wrong but we have seen that forensic interpretation is far richer and more informative than simple yes/no answers. In a source level proficiency test we expect the participants to respond with a statement of evidential weight in relation to one of two clearly stated propositions. Support thus expressed for a proposition that is, in fact, false is undesirable because it is misleading not “wrong”. Obviously, the desirable outcome of the proficiency test is a small value for the expected weight of evidence in relation to a false proposition. But whatever the outcome, the study must be seen as a learning exercise for all participants: the pool of knowledge has grown. The notion of an error rate to be presented to courts is misconceived because it fails to recognise that the science moves on as a result of proficiency tests. The work led by Found and Rogers [23] has shown how the profession of handwriting comparison in Australia and New Zealand has grown in stature because of the culture of advancing knowledge through repeated study under controlled conditions. To repeat then, our vision is not of the black box/error rate but of continuous development through calibration and feedback of opinions.

A striking example of forensic calibration is the evolution of fingerprints evidence from the identification paradigm to the logical paradigm via mathematical modelling [24,25]. Instead of the categorical identification, we have a mathematical approach that leads to a likelihood ratio. The validation of such approaches is founded on two desiderata: we require large likelihood ratios in cases in which the prosecution proposition is true; and small likelihood ratios in cases in which the defence proposition is true. Investigation of performance in relation to these two desiderata is undertaken by considering two sets of comparisons: one set in which it is known that the two samples came from the same source; and one set in which it is known that the two samples came from different sources. There have been major advances over recent years in how the likelihood ratio distributions from such experiments may be compared and evaluated (Ramos [26], Brümmer [27] see also Robertson et al. [28] for a layman’s introduction to calibration). The elegance and performance of such methods far transcends the crude PCAST notion of “false positive” and “false negative” error rates.

### 6.3. Knowledge and data

The PCAST report focuses on “feature comparison” methods and, as we have explained, this has meant that it is concerned with inference relating to source level propositions. At this level, the report sees data as the sole means for assigning probabilities. An important part of the role of the forensic scientist is concerned with inference with regard to activity level propositions. Consider, for example, a question of the form “what is the probability of finding this number of fragments of glass on Mr POI’s jacket if he is the person who smashed the window at the crime scene?” The answer is heavily dependent on circumstantial information (how large is the window? where was the person who smashed the window standing? was any implement used? how much time elapsed between the breaking of the window and the seizure of the jacket from Mr POI? etc.) and the variation in this between cases is vast. There is no single database to inform such probabilities. The scientist will, it is hoped, be thoroughly familiar with all of the published literature on glass transfer in crime cases [29] and may, if resources permit, carry out experiments that reproduce the current case circumstances. The knowledge and judgement of other scientists who have encountered similar questions is also relevant. We agree with PCAST that length of experience is not a measure of reliability of scientific opinion: the foundation is reliable knowledge. Too little effort has been devoted within the forensic sphere thus far to the harnessing of knowledge through knowledge based systems but see [29] for examples of how such a system was created for glass evidence interpretation.

We do not deny the importance of data collections but the view that data may replace judgement is misconceived. A data collection should be used to inform reliable knowledge not replace it.

We have explained that our view of the scientist is the antithesis of the PCAST “black box” automaton. Although there is a need for data, PCAST are mistaken in seeing it as the be all and end all: qualitative judgement will always be at the centre of forensic science evidence evaluation. We reject the PCAST vision of the scientist who gives a categorical opinion and a statement about the probability that the opinion is wrong. We see the model scientist as deeply knowledgeable about her domain of expertise and able to rationalise the opinion in terms that the jury will understand. The principles have been expressed elsewhere [11] as balance, logic, robustness and transparency. There is no place for the black box. We agree that the scientist should be able to provide the court with evidence of performance under controlled conditions. Found and Rogers [23] have provided a model for handwriting comparison



and we see such approaches as extending into other areas: the emphasis is on calibration of probabilistic assessments.

## 7. Conclusion

The 44th US president's request was "to consider whether there are additional steps that could usefully be taken on the scientific side to strengthen the forensic science disciplines and ensure the validity of forensic evidence used in the Nation's legal system" ([1], p. 1). We suggest that the report has very little emphasis on positive steps and does much to reinforce poor thinking and terminology.

Our own view of the future of forensic science is based on the principle that forensic inference should be founded on a logical framework for reasoning in the face of uncertainty. That framework is provided by probability theory coupled with the recognition that probability is necessarily subjective and conditioned by knowledge and judgement. It follows that our view of the forensic scientist is a knowledgeable, logical and reasonable person. Whereas data collections are valuable they should be viewed within the context of reliable knowledge. The overarching paradigm of reliable knowledge should be founded on the notion of knowledge management, including comprehensive systems for the calibration of expert opinion.

## References

- [1] President's Council of Advisors on Science and Technology, Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, Washington DC, 2016. [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensic\\_science\\_report\\_final.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf).
- [2] Federal Bureau of Investigation—FBI, Comments on: President's Council of Advisors on Science and Technology Report to the President on Forensic Science in Federal Criminal Courts: Ensuring Scientific Validity of Pattern Comparison Methods. September 20, 2016. [www.fbi.gov/file-repository/fbi-pcast-response.pdf/view](http://www.fbi.gov/file-repository/fbi-pcast-response.pdf/view).
- [3] National District Attorneys Association—NDAA, Report Entitled Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. November 16, 2016. <http://www.ciclt.net/ul/ndaajustice/PCAST/NDAAPCASTResponseFINAL.pdf>.
- [4] G.S. Morrison, D.H. Kaye, D.J. Balding, D. Taylor, P. Dawid, C.G.G. Aitken, S. Gittelson, G. Zadora, B. Robertson, S. Willis, S. Pope, M. Neil, K.A. Martire, A. Hepler, R.D. Gill, A. Jamieson, J. de Zoete, R.B. Ostrum, A. Caliebe, A comment on the PCAST report: skip the match/non-match stage, *Forensic Sci. Int.* 272 (2017) e7–e9, doi:<http://dx.doi.org/10.1016/j.forsciint.2016.10.018>.
- [5] Association of Firearm and Tool Mark Examiners AFTE, Response to PCAST Report on Forensic Science. October 31, 2016. [https://afte.org/uploads/documents/AFTE\\_PCAST\\_Response.pdf](https://afte.org/uploads/documents/AFTE_PCAST_Response.pdf).
- [6] Bureau of Alcohol Tobacco Firearms and Explosives—ATF, ATF Response to the President's Council of Advisors on Science and Technology Report. September 21, 2016. [https://www.theiai.org/president/20160921\\_ATF\\_PCAST\\_Response.pdf](https://www.theiai.org/president/20160921_ATF_PCAST_Response.pdf).
- [7] The International Association for Identification (IAI), IAI Response to the President's Council of Advisors on Science and Technology Report, 2016. [https://www.theiai.org/president/IAI\\_PCAST\\_Response.pdf](https://www.theiai.org/president/IAI_PCAST_Response.pdf).
- [8] President's Council of Advisors on Science and Technology, An addendum to the PCAST report on forensic science in criminal courts, Washington DC, 2017. [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensics\\_addendum\\_finalv2.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf).
- [9] C.G.G. Aitken, F. Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd ed, John Wiley & Sons Ltd., Chichester, 2004.
- [10] C. Aitken, P. Roberts, G. Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*, London, 2011. <http://www.rss.org.uk/uploadedfiles/userfiles/files/Aitken-Roberts-Jackson-Practitioner-Guide-1-WEB.pdf>.
- [11] Expressing evaluative opinions: a position statement, *Sci. Justice* 51 (1) (2011) 1–2, doi:<http://dx.doi.org/10.1016/j.scijus.2011.01.002>.
- [12] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A model for case assessment and interpretation, *Sci. Justice* 38 (3) (1998) 151–156, doi:[http://dx.doi.org/10.1016/S1355-0306\(98\)72099-4](http://dx.doi.org/10.1016/S1355-0306(98)72099-4).
- [13] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, *Sci. Justice* 38 (4) (1998) 231–240, doi:[http://dx.doi.org/10.1016/S1355-0306\(98\)72117-3](http://dx.doi.org/10.1016/S1355-0306(98)72117-3).
- [14] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, Case pre-assessment and review in a two-way transfer case, *Sci. Justice* 39 (2) (1999) 103–111, doi:[http://dx.doi.org/10.1016/S1355-0306\(99\)72028-9](http://dx.doi.org/10.1016/S1355-0306(99)72028-9).
- [15] P.L. Kirk, The ontogeny of criminalistics, *J. Crim. Law Criminol. Police Sci.* 54 (1963) 235–238.
- [16] D.A. Stoney, What made us ever think we could individualize using statistics, *J. Forensic Sci. Soc.* 31 (2) (1991) 197–199, doi:[http://dx.doi.org/10.1016/S0015-7368\(91\)73138-1](http://dx.doi.org/10.1016/S0015-7368(91)73138-1).
- [17] A. Biedermann, S. Bozza, F. Taroni, Decision theoretic properties of forensic identification: underlying logic and argumentative implications, *Forensic Sci. Int.* 177 (2–3) (2008) 120–132, doi:<http://dx.doi.org/10.1016/j.forsciint.2007.11.008>.
- [18] A. Biedermann, S. Bozza, F. Taroni, The decisionalization of individualization, *Forensic Sci. Int.* 266 (2016) 29–38, doi:<http://dx.doi.org/10.1016/j.forsciint.2016.04.029>.
- [19] W.C. Thompson, E.L. Schumann, Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy, *Law Hum. Behav.* 11 (3) (1987) 167–187, doi:<http://dx.doi.org/10.1007/BF01044641>.
- [20] E.H. Holder, M.L. Leary, J.H. Laub, DNA for the Defense Bar, U.S. Department of Justice Office of Justice Programs, Washington, DC, 2012.
- [21] National Research Council - Committee on DNA Technology in Forensic Science, *DNA Technology in Forensic Science*, National Academy Press, Washington, D.C, 1992.
- [22] I. Evett, The logical foundations of forensic science: towards reliable knowledge, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370 (1674) (2015), doi:<http://dx.doi.org/10.1098/rstb.2014.0263>.
- [23] B. Found, D. Rogers, The initial profiling trial of a program to characterize forensic handwriting examiners' skill, *J. Am. Society of Questioned Document Examiners* 6 (2) (2003) 72–81.
- [24] C. Champod, C.J. Lennard, P.A. Margot, M. Stoilovic, *Fingerprints and other Ridge Skin Impressions*, CRC Press, Boca Raton, 2016.
- [25] C. Neumann, I.W. Evett, J. Skerrett, Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, *J. Roy. Stat. Soc. Ser. A. (Stat. Soc.)* 175 (Part 2) (2012).
- [26] D. Ramos, J. Gonzalez-Rodriguez, G. Zadora, C. Aitken, Information-theoretical assessment of the performance of likelihood ratio computation methods, *J. Forensic Sci.* 58 (6) (2013) 1503–1518, doi:<http://dx.doi.org/10.1111/1556-4029.12233>.
- [27] N. Brümmer, J. du Preez, Application-independent evaluation of speaker detection, *Comput. Speech Language* 20 (2006) 230–275, doi:<http://dx.doi.org/10.1016/j.csl.2005.08.001>.
- [28] G.A. Robertson, C.E.H. Vignaux, *Interpreting Evidence—Evaluating Forensic Science in the Courtroom*, 2nd ed., John Wiley & Sons, Ltd., Chichester, 2016.
- [29] J.M. Curran, T.N. Hicks, J.S. Buckleton, *Forensic Interpretation of Glass Evidence*, CRC Press LLC, Boca Raton, 2000.

## Fwd: Rule 702 Subcommittee

---

**From:** "Antell, Kira M. (OLP)" <(b) (6)>  
**To:** "Hunt, Ted (ODAG)" <(b) (6)>  
**Date:** Tue, 05 Jun 2018 11:06:26 -0400  
**Attachment** national academy of science report pdf (3.63 MB); ATT00001.htm (216 bytes); pcast\_forensic\_science\_report\_final.pdf (1.86 MB); ATT00002.htm (216 bytes); Giannelli article on forensics.pdf (3.9 MB); ATT00003.htm (216 bytes); NEUROSCIENCE AND THE CIVILCRIMINAL DAUBERT DIVIDE.doc (303.09 kB); ATT00004.htm (216 bytes); TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM.doc (586.44 kB); ATT00005.htm (216 bytes); WILL HISTORY BE SERVED BY THE NAS REPORT ON FORENSIC SCIENCE AND THE ROLE OF THE.doc (389.53 kB); ATT00006.htm (216 bytes); drafting alternatives for rule 702 subcommittee.docx (18.76 kB); ATT00007.htm (168 bytes)

---

Sent from my iPhone

Begin forwarded message:

**From:** "Daniel Capra" <(b) (6)>  
**To:** "(b) (6) email for Thomas Schroeder" <(b) (6) email for Thomas Schroeder>, "Collin, Daniel" <(b) (6)>, "(b) (6) email for AJ Kramer" <(b) (6)>, "Shapiro, Elizabeth (CIV)" <(b) (6)>  
**Cc:** "Antell, Kira M. (OLP)" <(b) (6)>, "(b) (6) email for Debra Livingston" <(b) (6) email for Debra Livingston>  
**Subject:** Rule 702 Subcommittee

This is a "getting started" email for the Rule 702 Subcommittee appointed by Judge Livingston and chaired by Judge Schroeder. I am here to help the subcommittee's work in any way I can.

It is my understanding that the Subcommittee is to consider two basic issues:

What the Advisory Committee can/should do regarding forensics, which is subdivided into three questions:

- a. A possible rule amendment to regulate overstatement by experts (maybe all experts, maybe only forensic experts) --- together with a committee note that might speak more broadly about forensics.
- b. A more minor rule amendment, as a kind of coat hanger for an advisory committee note. That note might speak broadly about forensics and/or refer the reader to other sources, such as the FJC manual, NAS report, etc.
- c. Non-rule related ventures, such as working with the FJC on training programs and on the new Manual.

A possible amendment to Rule 702 directed mainly to civil cases, restoring the gatekeeping function on the questions of sufficiency of basis and reliability of application. This is in reaction to the many courts that have found these factors in Rule 702 to be question of weight and not admissibility

Judge Schroeder and I have conferred and we would like to have a conference call on the afternoon of July 11, to talk about how the Subcommittee can meet the goal

**Please let us know by email of available times you have that afternoon --- or if you are not available at all.**

I will also arrange to have Joe Cecil on the call, so he can give some more background about the FJC manual and training programs.

In the meantime, I am attaching a number of things for your reading pleasure:

The NAS Report, which most if not all of you have of course read, but I direct your attention to page 85-111. In that section, NAS arguably seems to criticize part of the 2000 Committee Note to Rule 702, which cites a case on handwriting and states that experience-based testimony can be reliable. Judge Livingston suggests that the Committee think about whether something can be done to address that passage in the Committee Note.

The PCAST report, which I attach only for the reference to Rule 702 and the Advisory Committee. In the section from pages 40-43, PCAST suggests that the Rule 702 Note actually is sufficient for courts to use to regulate forensic expert testimony. What PCAST suggests, in the recommendations section, is not a retroactive change in that Note, but rather a detailed best practices manual by way of an Advisory Committee Note.

A recent article by Professor Paul Giannelli on forensic evidence --- Professor Giannelli was a co-author of the current FJC manual chapter on forensics.

An article by Erin Murphy describing findings on the difference in the courts in applying Daubert in civil and criminal cases.

A recent note from the NYU Law review on how to resolve the problem of judicial deference to forensic evidence.

An article by Jane Moriarty on the asymmetry in application of Daubert in civil and criminal cases and how the NAS report might be used to address that.

A short memo by me, laying out drafting alternatives to address the two issues that the Subcommittee has on its agenda.

Finally, there is no specific deadline for the Subcommittee's work, but we would like to at least be able to report on progress at the October meeting.

Please let me know if you have any questions or comments. I look forward to working with you, and to talking with you on July 11.

Daniel J. Capra  
Reed Professor of Law  
Fordham Law School  
New York, New York

(b) (6)



**The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:**

**Document Title:           Strengthening Forensic Science in the United States: A Path Forward**

**Author:                     Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council**

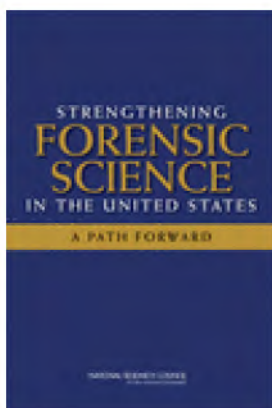
**Document No.:           228091**

**Date Received:           August 2009**

**Award Number:          2006-DN-BX-0001**

**This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.**

**Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.**



## Strengthening Forensic Science in the United States: A Path Forward

Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council

ISBN: 0-309-13131-6, 352 pages, 6 x 9, (2009)

This PDF is available from the National Academies Press at: <http://www.nap.edu/catalog/12589.html>

Visit the [National Academies Press](http://www.nap.edu) online, the authoritative source for all books from the [National Academy of Sciences](http://www.nap.edu), the [National Academy of Engineering](http://www.nap.edu), the [Institute of Medicine](http://www.nap.edu), and the [National Research Council](http://www.nap.edu):

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Explore our innovative research tools – try the “[Research Dashboard](#)” now!
- [Sign up](#) to be notified when new books are published
- Purchase printed books and selected PDF files

**Thank you for downloading this PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, [visit us online](#), or send an email to [feedback@nap.edu](mailto:feedback@nap.edu).**

**This book plus thousands more are available at <http://www.nap.edu>.**

Copyright © National Academy of Sciences. All rights reserved.

Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. [Request reprint permission for this book.](#)

THE NATIONAL ACADEMIES

*Advisers to the Nation on Science, Engineering, and Medicine*

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

September 19, 2017

**FORENSIC SCIENCE: *DAUBERT'S* FAILURE**

59 CASE W. RES. L. REV. \_\_ (revised 9/18/17)

Paul C. Giannelli  
Distinguish University Professor Emeritus  
Case Western Reserve University

- I. Introduction
  - A. *Daubert*/Rule 702
  - B. National Academy of Sciences Forensic Report (2009)
- II. Discredited Techniques
  - A. Bite Mark Comparisons
  - B. Microscopic Hair Analysis
  - C. Arson Evidence
  - D. Comparative Bullet Lead Analysis
- III. Misleadingly Presented Techniques
  - A. Firearms & Toolmark Identifications
  - B. Fingerprint Examinations
- IV. Forensic Science Research
  - A. National Commission on Forensic Science (2013-17)
  - B. White House PCAST Report (2016)
- V. Independent Scientific Review
- VI. Conclusion

*“The man who discovers a new scientific truth has previously had to smash to atoms almost everything he had learnt, and arrives at the new truth with hands bloodstained from the slaughter of a thousand platitudes.”* — Jose Oreta y Gasset, *The Revolt of the Masses* ch. XIV (1930).

## I. INTRODUCTION

In 2015, Judge Alex Kozinski of the Ninth Circuit Court of Appeals noted that “[m]any defendants have been convicted and spent countless years in prison based on evidence by arson experts who were later shown to be little better than witch doctors.”<sup>1</sup> In the same year, Dr. Jo Handelsman, a White House science advisor, observed: “Suggesting that bite marks [should] still be a seriously used technology is not based on science, on measurement, on something that has standards, but more of a gut-level reaction.”<sup>2</sup> According to Judge Catharine Easterly of the D.C. Court of Appeals, “[a]s matters currently stand, a certainty statement regarding toolmark pattern matching has the same probative value as the vision of a psychic.”<sup>3</sup> A *New York Times* editorial echoed these sentiments: “And the courts have only made the problem worse by purporting to be scientifically literate, and allowing in all kinds of evidence that would not make it within shouting distance of a peer-reviewed journal. Of the 329 exonerations based on DNA testing since 1989, more than one-quarter involved convictions based on ‘pattern’ evidence — like hair samples, ballistics, tire tracks, and bite marks — testified to by so-called experts.”<sup>4</sup>

---

<sup>1</sup> Alex Kozinski, *Criminal Law 2.0*, 44 GEO. L.J. ANN. REV. CRIM. PROC. iii, v (2015). See also *Almeciga v. Ctr. for Investigative Reporting, Inc.*, 185 F. Supp. 3d 401, 415 (S.D.N.Y. 2016) (“There have been too many pseudo-scientific disciplines that have since been exposed as profoundly flawed, unreliable, or baseless for any Court to take this [gate-keeping] role lightly.”).

<sup>2</sup> See Radley Balko, *A High-Ranking Obama Official Just Called for the “Eradication” of Bite Mark Evidence*, WASH. POST, July 22, 2015 (quoting remarks presented at the International Symposium on Forensic Science Error Management — Detection, Measurement and Mitigation, Arlington, Virginia (July 20–24, 2015), organized by the National Institute of Standards and Technology (NIST)).

<sup>3</sup> *Williams v. United States*, 130 A.3d 343, 355 (D.C. 2016) (Easterly, J., concurring).

<sup>4</sup> Editorial, *Junk Science at the F.B.I.*, N.Y. TIMES, Apr. 27, 2015. See also Eric S. Lander, *Fix the Flaws in Forensic Science*, N.Y. TIMES, Apr. 21, 2015 (“No expert should be permitted to testify without showing three things: a public database of patterns from many representative samples; precise and objective criteria for declaring matches; and peer reviewed published studies that validate the methods.”).



These criticisms are valid — which raises a puzzling and consequential question: Why didn't the Supreme Court's "junk science" decision, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,<sup>5</sup> prevent or restrict the admissibility of testimony based on flawed forensic techniques? *Daubert* was decided in 1993, nearly twenty-five years ago.

#### A. *Daubert/Rule 702*

*Daubert* was considered a revolutionary decision.<sup>6</sup> It "radically changed the standard for admissibility of scientific testimony"<sup>7</sup> by sweeping away the *Frye* "general acceptance" test,<sup>8</sup> which had been the majority rule in both federal and state cases.<sup>9</sup> The *Frye* standard gave great deference to the views of forensic practitioners and not to empirical testing.<sup>10</sup> *Daubert* promised to be different. The Supreme Court held that "[p]roposed testimony must be supported by appropriate validation — i.e., 'good grounds,' based on what is known. In short, the requirement that an expert's testimony pertain to 'scientific knowledge' establishes a standard of evidentiary reliability."<sup>11</sup> In making this reliability determination, the *Daubert* Court highlighted five factors: (1) testing, (2) peer review and publication, (3) error rate, (4) maintenance of standards, and (5) general acceptance. The first and *most important factor* is empirical testing. The other factors are supplementary. Peer review and publication are designed to expose defects in testing. Acceptance of a technique within the scientific community is achieved through the publication of valid test results. Similarly, both error rates and standards are derived from testing.

*Daubert* was followed in 1999 by *Kumho Tire v. Carmichael*,<sup>12</sup> which held that *Daubert*'s reliability standard applied to all expert testimony, not only

---

<sup>5</sup> 509 U.S. 579 (1993).

<sup>6</sup> See David L. Faigman, *Is Science Different for Lawyers?*, 297 SCIENCE 339, 340 (2002) ("*Daubert* initiated a scientific revolution in the law.>").

<sup>7</sup> *United States v. Barnette*, 211 F.3d 803, 815 (4th Cir. 2000). See also *United States v. Alatorre*, 222 F.3d 1098, 1100 (9th Cir. 2000) ("*Daubert* has become ubiquitous in federal trial courts.>").

<sup>8</sup> *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (stating that a technique "must be sufficiently established to have gained general acceptance in the particular field in which it belongs").

<sup>9</sup> See Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, A Half-Century Later*, 80 COLUM. L. REV. 1197 (1980).

<sup>10</sup> See Michael J. Saks, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 HASTINGS L.J. 1069, 1138 (1998) ("*Frye* does not work because its measure of validity is the judgment of 'the field,' and the field may consist of nonsense. For example, the *Frye* doctrine cannot exclude astrology.>").

<sup>11</sup> *Daubert*, 509 U.S. at 590 (emphasis added).

<sup>12</sup> 526 U.S. 137 (1999).

scientific evidence. By 2000, the Supreme Court was describing *Daubert* as establishing an “exacting” standard.<sup>13</sup> In the same year, Federal Rule of Evidence 702 was amended to incorporate the *Daubert/Kumho* standard.<sup>14</sup> Although a handful of jurisdictions continue to apply the *Frye* test, about forty jurisdictions have adopted the *Daubert* standard in one form or another.<sup>15</sup>

During this time, there was no shortage of commentary on the lack of empirical research in forensic science.<sup>16</sup> For example, shortly after *Daubert* was decided, Professor Margaret Berger wrote: “Considerable forensic evidence made its way into the courtroom without empirical validation of the underlying theory and/or its particular application.”<sup>17</sup> After *Kumho*, two commentators citing bite mark, hair, and firearm analysis observed that “little rigorous, systematic research has been done to validate the discipline’s basic premises and techniques, and in each area there was no evident reason why such research would be infeasible.”<sup>18</sup>

Notwithstanding *Daubert*’s promise, scholars soon discerned its uneven

---

<sup>13</sup> *Weisgram v. Marley Co.*, 528 U.S. 440, 455 (2000).

<sup>14</sup> After *Daubert*, the Court decided *General Elec. Co. v. Joiner*, 522 U.S. 136 (1997), which established the standard for appellate review (abuse of discretion) for applying the *Daubert* factors. *Daubert*, *Joiner*, and *Kumho* make up what is known as the *Daubert* Trilogy.

<sup>15</sup> See 1 PAUL C. GIANNELLI ET AL., *SCIENTIFIC EVIDENCE* § 1.06 (5th ed. 2012).

<sup>16</sup> A few perceptive scholars had noted the lack of empirical testing prior to *Daubert*. See Randolph N. Jonakait, *Forensic Science: The Need for Regulation*, 4 HARV. J. L. & TECH. 109, 137 (1991) (“Forensic science is supported by almost no research. The laboratory practices are based on intuitions and deductions, not on empirical proof.”); D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy For Rational Knowledge: The Lessons of Handwriting Identification “Expertise,”* 137 U. PA. L. REV. 731, 738 (1989) (“Our literature search for empirical evaluation of handwriting identification turned up one primitive and flawed validity study from nearly 50 years ago, one 1973 paper that raises the issue of consistency among examiners but presents only uncontrolled impressionistic and anecdotal information not qualifying as data in any rigorous sense, and a summary of one study in a 1978 government report. Beyond this, nothing.”); Michael J. Saks & Jonathan J. Koehler, *What DNA “Fingerprinting” Can Teach the Law About the Rest of Forensic Science*, 13 CARDOZO L. REV. 361, 372 (1991) (“[F]orensic scientists, like scientists in all other fields, should subject their claims to methodologically rigorous empirical tests. The results of these tests should be published and debated. Until such steps are taken, the strong claims of forensic scientists must be regarded with far more caution than they traditionally have been.”).

<sup>17</sup> Margaret A. Berger, *Procedural Paradigms for Applying the Daubert Test*, 78 MINN. L. REV. 1345, 1354 (1994) (“Courts never required some of the most venerable branches of forensic science — such as fingerprinting, ballistics, and handwriting — to demonstrate their ability to make unique identifications.”).

<sup>18</sup> Paul C. Giannelli & Edward J. Imwinkelried, *Scientific Evidence: The Fallout from the U.S. Supreme Court’s Decision in Kumho Tires*, 14 CRIMINAL JUSTICE, Winter 2000, at 12, 40. For an insightful analysis of how identification science was accepted by the courts, see Saks, *supra* note 10.



application in civil and criminal cases: “[T]he heightened standards of dependability imposed on expertise proffered in civil cases has continued to expand, but . . . expertise proffered by the prosecution in criminal cases has been largely insulated from any change in pre-*Daubert* standards or approach.”<sup>19</sup> The title of a 2005 article pretty much summed up the state of the law – “*The (Near) Irrelevance of Daubert to Criminal Justice*.”<sup>20</sup> In short, *Daubert*-lite.

## B. National Academy of Sciences Forensic Report (2009)

In 2006 Congress entered the picture by authorizing the National Academy of Sciences (NAS) to conduct a study of forensic science. After a three-year investigation, NAS issued a landmark report. One of its most riveting passages concluded: “Among existing forensic methods, only nuclear DNA analysis has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between an evidentiary sample and a specific individual or source.”<sup>21</sup> The report went on to state that “some forensic science disciplines are supported by little rigorous systematic research to validate the discipline’s basic premises and techniques.”<sup>22</sup> Such common forensic techniques as fingerprint examinations,<sup>23</sup> firearms (ballistics) and toolmark identifications,<sup>24</sup> handwriting examinations,<sup>25</sup> microscopic hair analysis,<sup>26</sup> and bite

---

<sup>19</sup> D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99, 149 (2000). In addition, an extensive study of reported criminal cases found that “the *Daubert* decision did not impact on the admission rates of expert testimony at either the trial or appellate court levels.” Jennifer Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 PSYCHOL. PUB. POL’Y & L. 339, 364 (2002).

<sup>20</sup> Peter J. Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggestions for Reform*, 95 AM. J. PUB. HEALTH S107 (2005).

<sup>21</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCES, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 100 (2009) [hereinafter NAS FORENSIC REPORT].

<sup>22</sup> *Id.* at 22. At another point, the report stated: “The simple reality is that the interpretation of forensic evidence is not always based on scientific studies to determine its validity. This is a serious problem.” *Id.* at 8. See also *id.* at 6 (“Often there are no standard protocols governing forensic practice in a given discipline. And, even when protocols are in place . . . , they often are vague and not enforced in any meaningful way.”).

<sup>23</sup> *Id.* at 144 (Research is needed “[t]o properly underpin the process of friction ridge [fingerprint] identification.”).

<sup>24</sup> *Id.* at 154 (“Sufficient studies [on firearms identification] have not been done to understand the reliability and repeatability of the methods.”).

<sup>25</sup> *Id.* at 166 (“The scientific basis for handwriting comparisons needs to be strengthened.”).

<sup>26</sup> *Id.* at 161 (“[T]estimony linking microscopic hair analysis with particular defendants is highly unreliable.”).

mark comparisons<sup>27</sup> fell into this category.

Not only did the NAS report highlight flaws in forensic science, it sharply criticized the judiciary for failing to demand the validation that *Daubert* required: “The bottom line is simple: In a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, *and the courts have been utterly ineffective in addressing this problem.*”<sup>28</sup> In a later passage, the report declared that “*Daubert* has done little to improve the use of forensic science evidence in criminal cases.”<sup>29</sup> The disparate treatment of civil actions and criminal prosecutions was also noted. After finding that “trial judges rarely exclude or restrict expert testimony offered by prosecutors,” the report commented: “[I]ronically, the appellate courts appear to be more willing to second-guess trial court judgments on the admissibility of purported scientific evidence in civil cases than in criminal cases.”<sup>30</sup>

Despite the NAS report, courts continued to admit the same evidence. Only a handful of courts applied the “exacting” standard that the Supreme Court said *Daubert* demanded.<sup>31</sup>

\* \* \*

In this article I examine the justice system’s failure by reviewing the status of two categories of forensic techniques. The first category involves discredited techniques: (1) bite mark analysis, (2) microscopic hair comparisons, (3) arson evidence, and (4) comparative bullet lead analysis. The second category involves techniques that have been misleadingly presented, which includes firearms/toolmark identifications and fingerprint examinations. Both categories present *Daubert* issues. I argue that the system’s failure can be traced back to its inability to demand and properly evaluate foundational research, i.e., *Daubert*’s first factor (empirical testing). Indeed, the justice system may be institutionally incapable of applying *Daubert* in criminal cases.

---

<sup>27</sup> *Id.* at 174 (“No thorough study has been conducted of large populations to establish the uniqueness of bite marks . . .”).

<sup>28</sup> *Id.* at 53 (emphasis added).

<sup>29</sup> *Id.* at 106.

<sup>30</sup> *Id.* at 11.

<sup>31</sup> *Weisgram*, 528 U.S. at 455. As former federal district judge Nancy Gertner noted: “[A] busy trial judge can rely on the decades of case law to legitimize decisions rejecting a hearing or motions in limine. And the trial judge can count on the Court of Appeals likely concluding that rejecting the challenge was not an abuse of the judge’s discretion.” Nancy Gertner, *Commentary on the Need for A Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 789, 790 (2011).



A different paradigm is needed, one that assigns an independent agency the responsibility of evaluating foundational research. As discussed in Part IV, this approach was recently recommended by the National Commission on Forensic Science (2013-17)<sup>32</sup> and the President’s Council of Advisors on Science and Technology (2016) (PCAST).<sup>33</sup> Both recommended that the National Institute of Standards and Technology (NIST) evaluate all forensic disciplines on a continuing basis, thereby injecting much needed scientific expertise into the criminal justice system. The recent reports on latent fingerprints<sup>34</sup> and arson investigations,<sup>35</sup> which were published by the American Association for the Advancement of Science (AAAS), buttress the need for independent scientific evaluations.

## II. DISCREDITED TECHNIQUES

### A. Bite Mark Comparisons

For decades, bite mark evidence has been admitted in hundreds of trials,<sup>36</sup> many of which were capital prosecutions.<sup>37</sup> No reported American case has rejected bite mark testimony. Moreover, it is not uncommon for courts to speak of bite mark comparison as a “science”<sup>38</sup> even an “exact science.”<sup>39</sup>

---

<sup>32</sup> In 2013, the Department of Justice (DOJ), in partnership with the National Institute of Standards and Technology (NIST), established the National Commission on Forensic Science to enhance the practice and improve the reliability of forensic science. The author served on the Commission.

<sup>33</sup> PRESIDENT’S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, REPORT TO THE PRESIDENT: FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016) [hereinafter WHITE HOUSE PCAST Report].

<sup>34</sup> See WILLIAM THOMPSON, ET AL., AM. ASSOC. ADVANCEMENT SCI. FORENSIC SCIENCE ASSESSMENTS: A QUALITY AND GAP ANALYSIS OF LATENT FINGERPRINT ANALYSIS (2017) [hereinafter AAAS FINGERPRINT REPORT]; .

<sup>35</sup> See JOSÉ ALMIRALI ET AL., AM. ASSOC. ADVANCEMENT SCI. FORENSIC SCIENCE ASSESSMENTS: A QUALITY AND GAP ANALYSIS: FIRE INVESTIGATION (2017) [hereinafter AAAS FIRE REPORT]

<sup>36</sup> See J. GIANNELI ET AL., *supra* note 15, § 13.05 (discussing the admissibility of bite mark evidence). In *Doyle v. State*, 263 S.W.2d 779 (Tex. Crim. App. 1954), a bite mark was left in a piece of cheese in a burglary case. Two decades later, in *Patterson v. State*, 509 S.W.2d 857, 862 (Tex. Crim. App. 1974), a prosecution expert matched the defendant’s teeth to a mark found on a murder victim.

<sup>37</sup> See Carrington Tucker, *Mississippi Innocence: The Convictions and Exonerations of Leven Brooks and Kennedy Brewer and the Failure of the American Promise*, 28 GEO. J. LEGAL ETHICS 123 (2015).

<sup>38</sup> See *People v. Marsh*, 441 N.W.2d 33, 35 (Mich. Ct. App. 1989) (“the science of bite mark analysis has been extensively reviewed in other jurisdictions”).

<sup>39</sup> See *State v. Sager*, 600 S.W.2d 541, 569 (Mo. Ct. App. 1980) (“an exact science”).

Acceptance of the technique is so deeply entrenched that some courts have taken judicial notice of its validity,<sup>40</sup> which means its reliability is indisputable.<sup>41</sup> Distinctive characteristics of a person's dentition were first used to identify skeletonized remains and individuals in mass disasters such as plane crashes.<sup>42</sup> Courts assumed that these distinctive characteristics can be transferred to another person's skin during a violent crime (e.g., homicides, rapes, and child abuse),<sup>43</sup> an assumption that overlooked some obvious problems. First, bite marks typically involve no more than the edges of six to eight front teeth, not thirty-two teeth with five anatomical surfaces that can be used when comparing a deceased person's dentition with X-rays. Second, bite marks do not reveal artifacts such as fillings, crowns, etc., all of which assist in associating human remains with a person's dental records.<sup>44</sup> Moreover, human skin is extremely malleable and thus subject to various types of distortion.<sup>45</sup> In addition, bite mark analysis is a subjective

---

<sup>40</sup> See *State v. Richards*, 804 P.2d 109, 112 (Ariz. 1990) (“[B]ite mark evidence is admissible without a preliminary determination of reliability . . .”); *People v. Middleton*, 429 N.E.2d 100, 101 (N.Y. 1981) (“The reliability of bite mark evidence as a means of identification is sufficiently established in the scientific community to make such evidence admissible in a criminal case, without separately establishing scientific reliability in each case . . .”); *State v. Armstrong*, 369 S.E.2d 870, 877 (W. Va. 1988) (judicially noticing the reliability of bite mark evidence).

<sup>41</sup> See FED. R. EVID. 201(b) (limiting judicial notice to a “fact that is not subject to reasonable dispute”).

<sup>42</sup> 1 GIANNELLI ET AL., *supra* note 15, at § 13.03 (discussing the admissibility of dental identifications).

<sup>43</sup> See *People v. Milone*, 356 N.E.2d 1350, 1358 (Ill. App. Ct. 1976) (“The concept of identifying a suspect by matching his dentition to a bite mark found at the scene of a crime is a logical extension of the accepted principle that each person's dentition is unique.”); *People v. Smith*, 443 N.Y.S.2d 551, 556-57 (Cty. Ct. 1981) (“The basic premise is the unique nature of individual dentition . . . and the virtually infinite number of individual bite configurations.”).

<sup>44</sup> “Restorations alone, with varying shapes, sizes, and restorative materials, may offer numerous points for comparison. In addition to restorations, the number of teeth, prostheses, decay, malposition, malrotation, peculiar shapes, root canal therapy, bone patterns, bite relationship, and oral pathology may all provide identifying characteristics.” 1 GIANNELLI ET AL., *supra* note 15, at 711.

<sup>45</sup> See I.A. Pretty & D. Sweet, *The Scientific Basis for Human Bitemark Analyses—A Critical Review*, 41 SCI. & JUST. 85, 87 (2001) (“Skin is a poor registration material since it is highly variable in terms of anatomical location, underlying musculature or fat, curvature, and looseness or adherence to underlying tissues. Skin is highly visco-elastic, which allows stretching to occur during either the biting process or when evidence is collected.”).

One study classified different types of distortion: Primary distortion occurs at the time of biting and results (1) from the dynamics of the biting process (dynamic distortion) and (2) from the features of the tissue bitten (tissue distortion). Secondary distortion occurs at a subsequent time. It can be subdivided into three categories. The first is time-related distortion, e.g., caused by subsequent healing or decomposition. Posture distortion results when the bite mark is viewed or recorded in a position that differs from the position at the time of biting. Photographic distortion results from the angle of the camera and the curvature of the body. See D.R. Sheasby & D.G. MacDonald, *A Forensic Classification of Distortion in Human Bite Marks*, 122 FORENSIC SCI. INT'L 75 (2001).



technique with no agreed-upon methodology.

### 1. Foundational Research

Despite overwhelming judicial approval, bite mark evidence is not supported by foundational research.<sup>46</sup> Indeed, the only rigorous studies are recent – and undercut the technique’s validity.<sup>47</sup> The 2009 NAS forensic report concluded that “the scientific basis is insufficient to conclude that bite mark comparisons can result in a conclusive match.”<sup>48</sup> Despite the NAS report, courts continued to permit expert testimony on the subject. For example, in *State v. Prade*,<sup>49</sup> decided in 2014, the expert testified that “bite mark evidence is generally accepted within the scientific community.”<sup>50</sup> Similarly, in *Coronado v. State*,<sup>51</sup> a different expert stated that he did not “agree with the NAS Report’s conclusion that bite mark analysis cannot result in a conclusive match” — adding “you do not have to be a ‘rocket scientist’ to see that, in some cases, there is a unique and distinct pattern of teeth that can be identified.”<sup>52</sup> In addition, these experts

---

<sup>46</sup> See Saks, *supra* note 10, at 1120 (“[R]ather than the field convincing the courts of the sufficiency of its knowledge and skills, admission by the courts apparently convinced the forensic odontology community that, despite their doubts, they really were able to perform bite mark identifications.”).

<sup>47</sup> Dr. Mary Bush and her colleagues at the Laboratory for Forensic Odontology, State University of New York at Buffalo, have published over a dozen studies that have undermined the assumptions underpinning bite mark evidence. See, e.g., Mary Bush et al., *Statistical Evidence for the Similarity of the Human Dentition*, 56 J. FORENSIC SCI. 118, 122 (2011) (“Our results show that given our measurement parameters, statements concerning dental uniqueness with respect to bitemark analysis in an open population are unsupportable. . . . Confidence in the notion of dental uniqueness in bitemark analysis has been based on anecdotal knowledge, the use of inappropriate statistics, and precedence of admission in the courtroom.”); Mary Bush et al., *Biomechanical Factors in Human Dermal Bitemarks in a Cadaver Model*, 54 J. FORENSIC SCI. 167 (2009) (23 bites were made in cadaver skin with the same dentition using an instrumented-biting machine. The cadavers were moved and re-photographed in different positions. Subsequent measurements showed differences between all bite marks. In addition, postural distortion was significant).

One survey of fifteen odontologists involved their opinions of six images of supposed bite marks. The “practitioner agreement was at best fair, with wide-ranging opinions on the origin, circumstance, and characteristics of the wound given for all six images.” M. Page et al., *Expert Interpretation of Bitemark Injuries — A Contemporary Qualitative Study*, 58 J. FORENSIC SCI. 664, 664 (2013).

<sup>48</sup> NAS FORENSIC REPORT, *supra* note 21, at 175.

<sup>49</sup> 9 N.E.3d 1072 (Ohio Ct. App. 2014).

<sup>50</sup> *Id.* at 1097.

<sup>51</sup> 384 S.W.3d 919 (Tex. Ct. App. 2012).

<sup>52</sup> *Id.* at 926.

rejected the valid research mentioned above<sup>51</sup> and both prosecutors and their experts attacked researchers without offering any foundational research.<sup>54</sup>

Unfortunately, the American Board of Forensic Odontology (ABFO) has fiercely defended bite mark analysis. To bolster its position, the ABFO conducted a study that was presented at a forensic conference in 2015.<sup>55</sup> As it turned out, the study undercut the ABFO's own position. Thirty-nine ABFO-certified bite mark experts — with an average of twenty years experience — examined 100 bite mark photographs. Each was asked three questions:

(1) Is there sufficient evidence in the presented materials to render an opinion on whether the patterned injury is a human bite mark?

(2) Is it a human bite mark, not a human bite mark, or suggestive of a human bite mark?

(3) Does the bite mark have distinct, identifiable arches and individual tooth marks?

The results to the first question were not reassuring. The thirty-nine experts agreed unanimously in only four out of the 100 cases. In only twenty cases was there 90 or more percent agreement. At the end of question two — whether the mark is a human bite mark — there were only sixteen cases with 90 or more percent agreement. At the end of the third question, there were only eight cases in which at least 90 percent of the analysts agreed.<sup>56</sup> Equally disturbing was the ABFO's decision to postpone publishing the results “until the organization can tweak the design of the study and conduct it again, a process that's expected to

---

<sup>51</sup> See *Prade*, 9 N.E.3d at 1098 (“As to Dr. Bush's cadaver studies, Dr. Wright testified that cadaver skin simply cannot compare with living skin. Dr. Wright explained that cadaver skin only distorts after a bite for two to three minutes at most because, unlike live skin, no bruising, contusions, or lacerations occur. Dr. Wright also testified that using a mechanical jaw to bite is problematic because the jaw operates on a fixed hinge that cannot mimic the wider range of movement that an actual jaw is capable of.”). But see I.A. Pretty & D. Sweet, *A Paradigm Shift in the Analysis of Bitemarks*, 201 *FORENSIC SCI. INT'L* 38, 40 (2010) (cadaver models have limitations but “there is little alternative for researchers to produce bitemarks of known origin”; use of anesthetized pigs to create *peri-mortem* injuries raises a different issue — i.e., differences between pigskin and human skin).

<sup>54</sup> See Radley Balko, *In Angry, Defensive Memo, Manhattan DA's Office Withdraws Bite Mark Evidence*, *WASH. POST.*, Jan. 13, 2016; Radley Balko, *Attack of the Bite Mark Matchers*, *WASH. POST.*, Feb. 18, 2015; Radley Balko, *The Path Forward on Bite Mark Matching — and the Rearview Mirror*, *WASH. POST.*, Feb. 20, 2015.

<sup>55</sup> The study is known as Construct Validity Bitemark Assessments Using the ABFO Bitemark Decision Tree (“Freeman/Pretty Study”).

<sup>56</sup> Radley Balko, *A Bite Mark Matching Advocacy Group Just Conducted a Study that Discredits Bite Mark Evidence*, *WASH. POST.*, Apr. 8, 2015.



take at least a year.”<sup>57</sup> In effect, a do-over. Meanwhile, an Associated Press analysis reported that at least twenty-four men convicted or charged with murder or rape based on bite marks have been exonerated since 2000.<sup>58</sup>

## 2. Texas Forensic Science Commission (2016)

Steven Chaney spent twenty-eight years in prison for murder based largely on bite mark evidence. When his conviction was overturned,<sup>59</sup> the Innocence Project filed a complaint on his behalf with the Texas Forensic Science Commission (TFSC).<sup>60</sup> In 2016, after a six-month investigation, the TFSC recommended a moratorium on the admission of bite mark testimony. It found that there is no scientific basis for claiming that a particular mark can be associated to a person’s dentition: “Any testimony describing human dentition as ‘like a fingerprint’ or incorporating similar analogies lacks scientific support.”<sup>61</sup> Similarly, “there is no scientific basis for assigning probability or statistical weight to an association, regardless of whether such probability or weight is expressed numerically (e.g., 1 in a million) or using some form of verbal scale (e.g., highly likely/unlikely).”<sup>62</sup>

TFSC was also alarmed that the ABFO study was not published due to “political and organizational pressures.” In the commission’s view, “such a resistance to publish scientific data contradicts the ethical and professional obligations of the profession as a whole, and is especially disconcerting when one considers the life and liberty interests at stake in criminal cases.”<sup>63</sup>

## 3. White House PCAST Report (2016)

---

<sup>57</sup> *Id.*

<sup>58</sup> See *Chaney v. State*, 775 S.W.2d 722, 725 (Tex. App. 1989) (A board-certified forensic odontologist “concluded that, in his opinion and with reasonable dental certainty, appellant made the bite marks on Swock’s body.”).

<sup>59</sup> Amanda Lee Myers, *Once Key in Some Cases, Bite-mark Evidence Now Derided as Unreliable*, Associated Press, Jun. 16, 2013.

<sup>60</sup> Texas created the Texas Forensic Science Commission (TFSC) in 2005 after a scandal required Houston to close its crime lab. See TEX. CRIM. PROC. CODE art. 38.01(4)(a)(3) (2005) (among other duties, the Commission should “investigate, in a timely manner, any allegation of professional negligence or misconduct that would substantially affect the integrity of the results of a forensic analysis conducted by an accredited laboratory, facility, or entity”). See generally Michael Hall, *False Impressions*, TEXAS MONTHLY, Jan. 2016.

<sup>61</sup> TEXAS FORENSIC SCI. COMM’N, FORENSIC BITEMARK COMPARISON COMPLAINT FILED BY NATIONAL INNOCENCE PROJECT ON BEHALF OF STEVEN MARK CHANEY – FINAL REPORT 11-12 (2016).

<sup>62</sup> *Id.* at 12.

<sup>63</sup> *Id.* at 13. See also Brandi Grissom, *Arguments Over Bitemarks Get Testy at Texas Forensic Science Commission Meeting*, DALLAS MORNING NEWS, Nov. 17, 2015.

In September 2016, the White House released its report on forensic science.<sup>64</sup> Regarding bite mark analysis, it concluded that (1) appropriately designed validation studies are lacking, (2) the few available studies had “very high” false-positive rates, (3) “inappropriate closed-set designs . . . are likely to underestimate the true false positive rate,” and (4) the studies show that experts “cannot even consistently agree on whether an injury is a human bite mark.”<sup>65</sup> Numerous cases support the last observation.<sup>66</sup>

\* \* \*

In sum, the courts have yet to reject bite mark evidence — a subjective method that is not supported by foundational research and lacks agreed-upon standards.<sup>67</sup> “Perhaps no discredited forensic assay has benefitted more from criminal courts’ abdication of gatekeeper responsibilities than bite mark analysis.”<sup>68</sup> Instead, it was the Innocence Project that spearheaded the challenges in this area, and in 2016 the Texas Forensic Science Commission became the first governmental body to seriously scrutinize the technique. Notwithstanding the NAS, PCAST, and TFSC reports, courts continue to admit bite mark evidence.<sup>69</sup>

---

<sup>64</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 9.

<sup>65</sup> “PCAST finds that bitemark analysis is far from meeting the scientific standards for foundational validity.” *Id.*

<sup>66</sup> See, e.g., *Ege v. Yukins*, 380 F. Supp. 2d 852, 878 (E.D. Mich. 2005) (“[T]he defense attempted to rebut Dr. Warnick’s testimony with the testimony of other experts who opined that the mark on the victim’s cheek was the result of *livor mortis* and was not a bite mark at all.”); *Czapleski v. Woodward*, 1991 U.S. Dist. Lexis 12567 (N.D. Cal. Aug. 30, 1991) (dentist’s initial report concluded that “bite” marks found on child were consistent with dental impressions of mother; several experts later established that the marks on child’s body were postmortem abrasion marks and not bite marks); *Kinney v. State*, 868 S.W.2d 463, 468 (Ark. 1994) (disagreement that marks were human bite marks); *People v. Noguera*, 842 P.2d 1160, 1165 n.1 (Cal. 1992) (“At trial, extensive testimony by forensic odontologists was presented by both sides, pro and con, as to whether the wounds were human bite marks and, if so, when they were inflicted.”); *State v. Duncan*, 802 So. 2d 533, 553 (La. 2001) (“Both defense experts testified that these marks on the victim’s body were not bite marks.”); *Stubbs v. State*, 845 So. 2d 656, 668 (Miss. 2003) (“Dr. Galvez denied the impressions found on Williams were the results of bite marks.”).

<sup>67</sup> See Michael J. Saks et al., *The Impending Death of Forensic Bitemark Identification*, 3 J. L. & BIOSCIENCES 1 (2016) (“[R]ecent reviews of the field’s claims, as well as recent empirical findings, have underscored the lack of reliability and validity of the most fundamental claims about the ability of forensic dentists to identify the source of bitemarks on human skin.”).

<sup>68</sup> M. Chris Fabricant & William Tucker Carrington, *The Shifted Paradigm: Forensic Science’s Overdue Evolution from Magic to Law*, 4 VA. J. CRIM. L. 1, 38 (2016).

<sup>69</sup> In *Commonwealth v. Ross*, No. CR 2038-2004, at 5 (C.P. Blair County, Pa., Mar. 8, 2017), the court admitted bite mark evidence, albeit limited, noting that “[t]he Commonwealth notes that no state or federal court has suppressed expert testimony in a criminal case based upon the NAS Report, and no courts have prohibited bite mark evidence based upon the PCAST or TFSC reports.” See also Radley Balko, *Incredibly, Prosecutors are Still Defending*



At the April 11, 2017 meeting of the National Commission on Forensic Science Keith Harward described how bite mark evidence resulted in his thirty-three years of imprisonment before he was exonerated by DNA evidence.<sup>70</sup> Incredibly, the next day the chairman of the National District Attorneys Association stated that his organization believes that bite mark evidence is a “reliable science.”<sup>71</sup>

## B. Microscopic Hair Analysis

In this examination, samples are first examined to identify features visible to the naked eye such as color and form, i.e., whether it is straight, wavy, or curved. Next, the sample is viewed microscopically to determine characteristics such as shaft form, hair diameter, and pigment size.

Experts have long acknowledged that a positive identification is not possible with microscopic hair analysis. Instead, examiners testify that a crime scene exemplar was “consistent with” a hair sample from the defendant. The probative value of this conclusion would, of course, vary if only a hundred people had microscopically indistinguishable hair as opposed to several million. Due to a lack of research, no one knows whether the crime scene hair could have come from 10 other persons or 100, 10,000, and so forth.<sup>72</sup> This important qualifying information was often omitted from the experts’ testimony, thus making marginal evidence appear misleadingly convincing.<sup>73</sup>

---

*Bite Mark Evidence*, WASH. POST, Jan 30, 2017.

<sup>70</sup> See Frank Green, *DNA Proves Man Innocent of 1982 Rape and Murder in Famous ‘Bite-mark’ Case, Lawyers Say*, RICHMOND TIMES-DISPATCH, Mar. 12, 2016 (Keith Harward case); Spencer S. Hsu, *Va. Exoneration Underscores Mounting Challenges to Bite-Mark Evidence*, WASH. POST, Apr. 8, 2016 (Keith Harward case).

<sup>71</sup> Penn Levy, *Sessions’ New Forensic Science Adviser Has a History of Opposing Pro Science Reforms*, MOTHER JONES, Aug. 10, 2017.

<sup>72</sup> As one hair examiner wrote: “If a pubic hair from the scene of a crime is found to be similar to those from a known source, [the courts] do not know whether the chances that it could have originated from another source are one in two or one in a billion.” B.D. Gaudette, *Probabilities and Human Pubic Hair Comparisons*, 21 J. FORENSIC SCI. 514, 514 (1976).

<sup>73</sup> Professor Berger explained the problem:

We allow eyewitnesses to testify that the person fleeing the scene wore a yellow jacket and permit proof that a defendant owned a yellow jacket without establishing the background rate of yellow jackets in the community. Jurors understand, however, that others than the accused own yellow jackets. When experts testify about samples matching in every respect, the jurors may be oblivious to the probability concerns if no background rate is offered, or may be unduly prejudiced or confused if the probability of a match is confused with the probability of guilt, or if a background rate is offered that does not have an adequate scientific foundation.

Berger, *supra* note 17, at 1357.



However, experts frequently went way beyond the “consistent with” language in their testimony, suggesting a rare association. For example, in the Edward Honaker case, the expert testified that the crime scene hair sample “was unlikely to match anyone” other than the defendant.<sup>74</sup> Honaker spent ten years in prison before DNA proved him innocent.<sup>75</sup> In another case, an expert testified that hair samples were “consistent microscopically” but then elaborated: “In other words, hairs are not an absolute identification, but they either came from this individual or there is — could be another individual *somewhere in the world* that would have the same characteristics to their hair.”<sup>76</sup> This is an implicit (and extreme) probability statement that lacks any empirical support.

Although microscopic hair analysis had long been judicially accepted,<sup>77</sup> its validity was suspect.<sup>78</sup> In 1995, a federal district court in *Williamson v. Reynolds* observed: “Although the hair expert may have followed procedures accepted in the community of hair experts, the human hair comparison results in this case were, nonetheless, scientifically unreliable.”<sup>79</sup> The court also noted that the “expert did not explain which of the ‘approximately’ 25 characteristics were consistent, any standards for determining whether the samples were consistent, how many persons could be expected to share this same combination of characteristics, or how he arrived at his conclusions.”<sup>80</sup> Williamson, who was five days from execution when he obtained habeas relief, was subsequently exonerated by DNA testing.<sup>81</sup>

---

<sup>74</sup> EDWARD CONNORS ET AL., CONVICTED BY JURIES, EXONERATED BY SCIENCE: CASE STUDIES IN THE USE OF DNA EVIDENCE TO ESTABLISH INNOCENCE AFTER TRIAL 58 (1996).

<sup>75</sup> *Id.*

<sup>76</sup> *Williamson v. Reynolds*, 904 F. Supp. 1529, 1554 (E.D. Okl. 1995) (emphasis added), *rev'd on this issue sub nom.*, *Williamson v. Ward*, 110 F.3d 1508, 1523 (10th Cir. 1997) (holding that due process, not *Daubert*, controls in federal habeas review).

<sup>77</sup> See Edward J. Imwinkelried, *Forensic Hair Analysis: The Case Against the Underemployment of Scientific Evidence*, 39 WASH. & LEE L. REV. 41, 62 (1982) (stating that “[t]he massive body of case law, liberally admitting even hair evidence of low probative value, dwarfs the handful of cases excluding hair evidence”).

<sup>78</sup> See Clive A. Stafford Smith & Patrick D. Goodman, *Forensic Hair Comparison Analysis: Nineteenth Century Science or Twentieth Century Snake Oil?*, 27 COLUM. HUM. RTS. L. REV. 227, 231 (1996) (“If the purveyors of this dubious science cannot do a better job of validating hair analysis than they have done so far, forensic hair comparison analysis should be excluded altogether from criminal trials.”).

<sup>79</sup> *Williamson*, 904 F. Supp. at 1558.

<sup>80</sup> *Id.* at 1554.

<sup>81</sup> See BARRY SCHECK ET AL., ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED 146 (2000) (noting that the hair evidence was shown to be “patently unreliable”). See also JOHN GRISHAM, *THE INNOCENT MAN: MURDER AND INJUSTICE IN A SMALL TOWN* (2006) (examining Williamson’s trial).

The *Williamson* opinion — perhaps the only thorough judicial analysis of microscopic hair comparisons — was all but ignored by other courts. In *Johnson v. Commonwealth*<sup>82</sup> (1999), the Kentucky Supreme Court upheld the admissibility of hair evidence, taking “judicial notice” of its reliability<sup>83</sup> and thus implicitly finding its validity indisputable.<sup>84</sup> Other courts echoed *Johnson*, not *Williamson*.<sup>85</sup> Indeed, ten years after *Williamson* was decided, a 2005 decision by the Connecticut Supreme Court observed (correctly) that “[t]he overwhelming majority of courts have deemed such evidence admissible.”<sup>86</sup>

Once again, the courts abdicated their responsibility. Indeed, hair evidence only began to be carefully scrutinized after a startling number of DNA exonerations were reported.<sup>87</sup> A 2008 study of 200 DNA exonerations found that expert testimony (55 percent) was the second leading type of evidence — after eyewitness identifications (79 percent) — used in wrongful conviction cases.<sup>88</sup> A subsequent investigation of trial transcripts underscored the role of hair analysis in the exoneration cases: “Of the 65 cases involving microscopic hair comparison in which transcripts were located, 25 cases, or 38%, had invalid forensic science testimony.”<sup>89</sup> The 2009 NAS report observed that “testimony linking microscopic hair analysis with particular defendants is highly unreliable.”<sup>90</sup>

### 1. FBI Hair Review

In May 2013, the Mississippi Supreme Court, in a 5-to-4 decision, rejected Willie Jerome Manning’s request for a stay of execution to permit DNA testing “potentially setting up what experts said would be a rare case in recent years in

---

<sup>82</sup> 12 S.W.3d 258 (Ky. 1999).

<sup>83</sup> *Id.* at 267.

<sup>84</sup> See FED. R. EVID. 201(b) (limiting judicial notice to a “fact that is not subject to reasonable dispute”).

<sup>85</sup> See 2 GIANNELLI ET AL., *supra* note 15, § 24.03, at 825 (noting the “limited impact of *Daubert*”).

<sup>86</sup> *State v. West*, 877 A.2d 787, 808 (Conn. 2005).

<sup>87</sup> In 1998, a Canadian judicial inquiry into the wrongful conviction of Guy Paul Morin was released. Morin’s original conviction was based, in part, on hair evidence. The judge conducting the inquiry recommended that “[c]riminal judges should undertake a more critical analysis of the admissibility of hair comparison evidence as circumstantial evidence of guilt.” HON. FRED KAUFMAN, THE COMMISSION ON PROCEEDINGS INVOLVING GUY PAUL MORIN (Ontario Ministry of the Attorney General 1998) (Recommendation 2). See also EDWARD CONNORS ET AL., *supra* note 74, 58 (listing cases).

<sup>88</sup> Brandon L. Garrett, *Judging Innocence*, 108 COLUM. L. REV. 55, 81 (2008).

<sup>89</sup> Brandon L. Garrett & Peter Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1, 14-15 (2009).

<sup>90</sup> NAS FORENSIC REPORT, *supra* note 21, at 161.



which a person is put to death with such requests unmet.”<sup>91</sup> A week later, the court unexpectedly stayed Manning’s execution — after the Department of Justice (DOJ) notified state officials that FBI experts had presented misleading testimony at his trial, including hair and firearms evidence.<sup>92</sup>

Soon after, the DOJ announced that Manning was but one of 120 cases — including twenty-seven death penalty prosecutions — in which improper microscopic hair analysis had been introduced in evidence.<sup>93</sup> For example, examiners claimed to connect a hair sample to a single person “to the exclusion of all others” or stated or suggested a probability for such a match from past casework.<sup>94</sup> The FBI review came after three District of Columbia men, who had been convicted of rape or murder in the early 1980s, were exonerated through DNA testing.<sup>95</sup> In one of these cases, the FBI expert testified: “Chances that it came from someone else were ‘one in 10 million.’”<sup>96</sup>

After further investigation, DOJ reported in 2015 that “FBI examiners had provided scientifically invalid testimony in more than 95 percent of cases where that testimony was used to inculpate a defendant at trial.”<sup>97</sup> *Commonwealth v. Perrott*<sup>98</sup> was one of the first cases to reach the courtroom as a consequence of the DOJ review. A superior court granted Perrott a new trial in 2016, criticizing the misleading use of hair evidence. The court noted: “In discussing the ‘microscopic characteristics’ of hair, [the expert] stated that these characteristics

---

<sup>91</sup> Campell Robertson, *Mississippi Inmate’s Bid for DNA Tests Is Denied With Tuesday Execution Set*, N.Y. TIMES, May 4, 2013, at A11.

<sup>92</sup> See Campell Robertson, *With Hours Left to Go, Execution Is Postponed*, N.Y. TIMES, May 8, 2013 (noting that the Department of Justice “disavow[ed] the degree of certainty expressed by F.B.I. forensic experts at the man’s trial”). See also Andrew Cohen, *A Ghost of Mississippi: The Willie Manning Capital Case*, THE ATLANTIC, May 2, 2013.

<sup>93</sup> See Jack Nicas, *Flawed Evidence Under a Microscope: Disputed Forensic Techniques Draw Fresh Scrutiny: FBI Says It Is Reviewing Thousands of Convictions*, WALL ST. J., July 18, 2013.

<sup>94</sup> Spencer S. Hsu, *U.S. Reviewing 27 Death Penalty Convictions for FBI Forensic Testimony Errors*, WASH. POST, July 17, 2013 (“[O]n the witness stand, several agents for years went beyond the science and testified that their hair analysis was a near-certain match.”).

<sup>95</sup> See Editorial, *Failures at the FBI Crime Lab*, WASH. POST, Apr. 20, 2012 (“Kirk L. Odom was incarcerated for 20 years and Donald E. Gates for nearly 30 for crimes they did not commit. Santae A. Tribble spent 28 years behind bars, even though DNA evidence now shows he almost undoubtedly was not the culprit.”).

<sup>96</sup> Martin Enserink, *Evidence on Trial*, 351 SCIENCE 1129, 1129, Mar. 11, 2016.

<sup>97</sup> WHITE HOUSE PCAST REPORT, *supra* note 33 at 3. See also Editorial, *Junk Science at the F.B.I.*, N.Y. TIMES, Apr. 27, 2015 (“a sweeping post-conviction review of 2,500 cases”); Hugh B. Kaplan, *DOJ Examiners Gave Bad Testimony in 90 Percent of Hair Comparison Cases*, BNA CRIM. L. REP. 77, Apr. 22, 2015.

<sup>98</sup> Nos. 85–5415, 5416, 5418, 5420, 5425, 2016 WL 380123 (Mass. Super. Ct. Jan. 26, 2016).

'make that hair somewhat unique.' He likened the 'subtle' characteristics of hair that 'make it somewhat unique' to the subtle differences in a human face.<sup>99</sup>

## 2. *White House PCAST Report (2016)*

In June 2016, the Department of Justice released proposed guidelines concerning hair testimony. Documentation purporting to support the validity and reliability of hair evidence accompanied the guidelines.<sup>100</sup> Listing several studies, the FBI concluded:

Based on these and other published studies, microscopic hair comparison has been demonstrated to be a valid and reliable scientific methodology. These studies have also shown that microscopic hair comparisons alone cannot lead to personal identification and it is crucial that this limitation be conveyed both in the written report and in testimony.<sup>101</sup>

The White House PCAST report, however, challenged the supporting documentation, which discussed only a handful of studies from the 1970s and 1980s but did not comment on subsequent studies that found "substantial flaws in the methodology and results of the key papers."<sup>102</sup> Moreover, "PCAST's own review of the cited papers [found] that these studies do not establish the foundational validity and reliability of hair analysis."<sup>103</sup>

---

<sup>99</sup> *Id.* at \*32. The expert also "asserted that the hairs 'matched' and showed a 'strong association.' In discussing the chance that the hair found on the victim's bed came from someone other than Perrot, [the expert] conceded the possibility, adding that during his ten years of experience 'it's extremely rare that I will have known hair samples from two different people that I can't tell apart.' [The expert] made these statements of confidence, despite being unable to recall at trial the length or diameter of the one hair found on the bed." *Id.*

<sup>100</sup> Office of Public Affairs, Justice Department Issues Draft Guidance Regarding Expert Testimony and Lab Reports in Forensic Science, June 3, 2016. These documents are known as the Uniform Language for Testimony and Reports. <https://www.justice.gov/opa/pr/justice-department-issues-draft-guidance-regarding>.

<sup>101</sup> Supporting Documentation for Department of Justice Proposed Uniform Language for Testimony and Reports for the Forensic Hair Examination Discipline at 4. [www.justice.gov/dag/file/877741/download](http://www.justice.gov/dag/file/877741/download).

<sup>102</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 13.

<sup>103</sup> *Id.* DOJ's supporting documents cited M.M. Houck & B. Budowle, *Correlation of Microscopic and Mitochondrial DNA Hair Comparisons*, 47 J. FORENSIC SCI. 964 (2002). This FBI study used mitochondrial DNA analysis to re-examine samples from previous FBI microscopic hair examination cases. The PCAST report did not accept that this study supported validity and reliability because the study showed that in 9 of 80 cases (11 percent) the microscopic examination found the hair indistinguishable but DNA analysis showed that the hairs came from different individuals.



\* \* \*

The bottom line, again, is the judiciary's dereliction in failing to curb the misuse of hair microscopy testimony. The Innocence Project's track record of DNA exonerations brought this issue to the fore. Indeed, the three exonerations in the District of Columbia triggered the FBI review. Yet, DOJ's proposed guidelines were based on "foundational research" that PCAST questioned.

### C. Arson Investigations

For decades arson investigators came from the "old school" of investigators — those who used intuition and a number of rules of thumb to determine whether a fire was incendiary. Critics of this approach complained that it lacked a scientific foundation. Rather, it was based on folklore that had been passed down from generation to generation — without any empirical testing.<sup>104</sup> As early as 1977, a government report noted that common arson indicators had "received little or no scientific testing" and "[t]here appears to be no published material in the scientific literature to substantiate their validity."<sup>105</sup> Through the 1980s, proponents of a science-based approach to arson investigations waged an uphill battle, finally winning a major victory in 1992 when the National Fire Protection Association (NFPA) published its *Guide for Fire and Explosion Investigations* (NFPA 921).<sup>106</sup>

#### 1. *Willingham Case*

Although NFPA 921 would subsequently become the bible in arson investigations,<sup>107</sup> it was published weeks after Cameron Todd Willingham was convicted for the arson-murders of his young children. Willingham, who was executed twelve years later, is the poster-boy for junk science in arson

---

<sup>104</sup> See JOHN J. LENTINI, SCIENTIFIC PROTOCOLS FOR FIRE INVESTIGATION ch. 8 (2006) (discussing myths of arson investigations).

<sup>105</sup> J.F. BOUDREAU ET AL., NATIONAL INSTITUTE OF LAW ENFORCEMENT AND CRIMINAL JUSTICE, LAW ENFORCEMENT ASSISTANCE ADMINISTRATION, U.S. DEPT OF JUSTICE, ARSON AND ARSON INVESTIGATION: A SURVEY AND ASSESSMENT (1977).

<sup>106</sup> NATIONAL FIRE PROTECTION ASSOCIATION, GUIDE FOR FIRE AND EXPLOSION INVESTIGATION (1992) [hereinafter NFPA 921]. The NFPA promotes fire prevention and safety. The most recent edition of NFPA 921 was published in 2011.

<sup>107</sup> See *United States v. Hebshtic* 754 F. Supp. 2d 89, 111 n.39 (D. Mass. 2010) (NFPA 921 "is widely accepted as the standard guide in the field of fire investigation."); Thomas M. May, *Fire Pattern Analysis, Junk Science, Old Wives Tales, and Ipse Dixit: Emerging Forensic 3D Imaging Technologies to the Rescue?*, 16 RICHMOND J.L. & TECH. 1, 5 (2010) (noting that NFPA 921 has "become the *de facto* national standard for fire scene examination and analysis").

investigations.<sup>108</sup>

At trial Deputy Fire Marshall Vasquez testified that “[t]he fire tells a story. I am just the interpreter. . . . And the fire does not lie. It tells me the truth.”<sup>109</sup> He told the jury that he had found twenty arson so-called “indicators” during his post-fire investigation of Willingham’s house.<sup>110</sup> One indicator was a low burning fire.<sup>111</sup> “All fire goes up,” Vasquez testified.<sup>112</sup> Thus, burn patterns on the lower walls and floor suggested that an accelerant was used.<sup>113</sup> This common-sense notion, however, has its limitations, especially when a fire occurs in a contained area, such as a house with its windows shut. Due to buoyancy, a thermal plume initially rises once a fire is ignited. As the fire continues, the plume reaches the ceiling, which causes it to spread outward towards the walls. When it reaches the walls, the combustion products press down from the ceiling creating an upper level, which continues to increase in depth and temperature. Eventually thermal radiation replaces convection as the principal method of heat transfer. At this point, every combustible surface in the room will spontaneously burst into flames. This transition phenomenon, known as the onset of “flashover,” can occur within minutes. After flashover, the entire room is burning, including the lower walls and floor. Flashover, according to one authority, is the point at which the fire transitions from a “fire in a room” to a “room on fire.”<sup>114</sup> At trial, prosecution

---

<sup>108</sup> See *Frontline: Death by Fire* (PBS television broadcast Oct. 19, 2010); David Grann, *Trial by Fire: Did Texas Execute an Innocent Man?*, *NEW YORKER*, Sep. 7, 2009, at 63; Michael Hall, *False Impressions*, *TEX. MONTHLY*, Jan. 2016, at 7 (“The 893-page report, released in April 2011, was anticlimactic for people looking for proof that Texas had executed an innocent man.”); Steve Mills & Maurice Possley, *Texas Man Executed on Disproved Forensics: Fire that Killed His 3 Children Could Have Been Accidental*, *CITR. TRIB.*, Dec. 9, 2004, at C1 (“Arson investigators in Texas have relied on old wives’ tales and junk science to send men to prison, and perhaps even the death chamber, top experts on fire behavior say.”).

<sup>109</sup> Transcript, *State v. Willingham*, No. 24240-CR (13th Dist., Tex. 1991), vol. XI, at 244 [hereinafter *Willingham transcript*], *aff’d*, *Willingham v. State*, 897 S.W.2d 351, 354 (Tex. Crim. App. 1995).

<sup>110</sup> A second expert’s testimony essentially tracked Vasquez’s.

<sup>111</sup> Vasquez testified that there was “char burning, like, for example, this is the bottom here. It’s burned down here at the bottom. That is an indicator in my investigation of an origin of fire because it’s the lowest part of the fire.” *Willingham transcript*, *supra* note 109, vol. XI, at 239. See also *Willingham*, 897 S.W.2d at 354 (“An expert witness for the State testified that the floors, front threshold, and front concrete porch were burned, which only occurs when an accelerant has been used to purposely burn these areas. This witness further testified that this igniting of the floors and thresholds is typically employed to impede firemen in their rescue attempts.”).

<sup>112</sup> *Willingham transcript*, *supra* note 109, vol. XI, at 232.

<sup>113</sup> “So when I found that the floor is hotter than the ceiling, that’s backwards, upside down. It shouldn’t be like that. The only reason that the floor is hotter is because there was an accelerant.” *Id.* at 256.

<sup>114</sup> LENTINI, *supra* note 104, at 68-70.



witnesses acknowledged that there was an explosion.<sup>115</sup> Consequently, a low burning fire is not necessarily indicative of an incendiary origin.

Moreover, some of Vasquez's other "indicators" — splotchy looking areas that he called "puddle configurations" and "pour patterns" — are present after flashover in an accidental fire.<sup>116</sup> Similarly, additional indicators, such as alligatoring (large shiny charred blisters on burned wood), are also explained by flashover. This phenomenon also accounts for another fact that Vasquez thought incriminatory. Willingham told investigators that he had attempted to save his daughters, but the heat was too great and he was forced to run from the house without shoes. Willingham's feet were not burned, and in Vasquez's mind, the burn debris on the floor made that impossible.<sup>117</sup> However, if Willingham left his home before flashover, his feet would not have been burned.

Charring under an aluminum threshold of an interior door provided still another clue. Here, again, this may occur in a flashover. Other perceived indicators — melted bed springs,<sup>118</sup> multiple points of origins,<sup>119</sup> and brown stains

---

<sup>115</sup> See Willingham transcript, *supra* note 109, vol. XI, at 75 ("The windows, the electricity started crackling and popping, and the top of the well — well, I was facing the side of the house, and it just blew out. The flames just blew out. . . . All the windows and the front room was engulfed.") (testimony of Dianne Barbe); *id.* at 96 ("We was running towards the house, me and my mother, we was fixing to go and try to get in, and that's when it was an explosion.") (testimony of Dianne Barbe). Vasquez mentioned flashover in his testimony (*id.* vol. XII, at 47-48), but he does not appear to understand its implications.

<sup>116</sup> According to Vasquez, a burn trailer was etched on the floor. Willingham transcript, *supra* note 109, vol. XI, at 244 ("You can see that on the burnt patterns on this puddle configuration on Exhibit No. 36. This is a strong indicator of a liquid.").

<sup>117</sup> "There was fire on the floor. . . . He had no injuries on his feet." *Id.* at 267.

<sup>118</sup> "[T]he springs were burned from underneath. This indicates there was a fire under this bed because of the burn underneath the bed." *Id.* at 241.

<sup>119</sup> "Multiple areas of origin indicate — especially if there is no connecting path, that they were intentionally set by human hands." Willingham transcript, *supra* note 109, at 255. There are two problems here. First, the fire scene did not exhibit multiple origins, according to independent experts. DOUGLAS CARPENTER ET AL., REPORT ON THE PEER REVIEW OF THE EXPERT TESTIMONY IN THE CASES OF STATE OF TEXAS V. CAMERON TODD WILLINGHAM AND STATE OF TEXAS V. EARNEST RAY WILLIS 11-12 (2006). Second, even if the fire scene had shown multiple points of origin, this would not necessarily indicate an intentional fire. LENTINI, *supra* note 104, at 461-62.



on a concrete floor<sup>120</sup> were also consistent with an accidental blaze.<sup>121</sup> Vasquez also relied on the presence of “crazed glass,” which are spider-web patterns on the windows as an indication of arson.<sup>122</sup> It was long believed that crazed glass resulted from a fire that burned fast and hot — i.e., one fueled by a liquid accelerant. Yet, subsequent research demonstrated that crazing occurs from rapid cooling when water from fire hoses is sprayed on heated windows.<sup>123</sup>

In retrospect, the most damning piece of evidence involved one of the numerous debris samples submitted for laboratory analysis.<sup>124</sup> It came from an area near the front door and was the *only* sample that tested positive for a chemical commonly used in charcoal lighter fluids. Nevertheless, this finding can be explained by the fact that a charcoal grill and lighter fluid were on the front porch at the time of the fire.<sup>125</sup> In fact, the negative results from the other samples supported Willingham’s case.<sup>126</sup>

Numerous nationally-recognized experts reviewed the arson testimony presented at Willingham’s trial and found it seriously flawed. The first examination of the record by an independent expert was submitted to the governor and the Board of Pardons and Parole days before Willingham’s execution. It

---

<sup>120</sup> Willingham transcript, *supra* note 109, vol. XI, at 248-49. Fire experts reviewing the evidence from Willingham’s trial pointed out that “[t]he behavior of concrete in fires, including the development of various colors, has been extensively studied.” CARPENTER ET AL., *supra* note 119, at 18. These experts concluded that there is simply “no scientific basis for Mr. Vasquez’s statement about the brown discoloration being an indication of the presence of accelerants.” *Id.*

<sup>121</sup> Vasquez’s testimony also demonstrated other misconceptions. A common one is that arson fires burn hotter and faster than “normal” fires: “You know, it makes the fire hotter. It’s not a normal fire.” Willingham transcript, *supra* note 109, vol. XI, at 249. However, the temperature of burning wood and burning gasoline are nearly identical, so to claim that a fire using liquid accelerants burns “hotter” than a wood fire is wrong. LENTINI, *supra* note 254, at 465.

<sup>122</sup> “The pieces of broken window glass on the ledge of the north windows to the northeast bedroom disclosed a crazed ‘spider webbing’ condition. This condition is an indication that the fire burned fast and hot.” CARPENTER ET AL., *supra* note 119, at 18 (citing Vasquez’s written report on the Willingham fire at 4).

<sup>123</sup> LENTINI, *supra* note 104, at 439 (“It is unclear why anyone ever thought that crazing of glass indicated rapid heating.”).

<sup>124</sup> In closing argument, the defense counsel referred to a “dozen samples.” Willingham transcript (vol. XIII), *supra* note 109, at 20.

<sup>125</sup> *Id.* at 15 (although photographs show a grill, Vasquez apparently did not know of the grill’s presence); *id.* at 16 (acknowledging that a fire-damaged charcoal lighter fluid container was found on the front porch).

<sup>126</sup> The prosecutor would later say that he “‘never did understand why they weren’t able to recover’ positive tests in these parts.” Grann, *supra* note 108, at 61. At trial, he argued that the “liquid burned away in that destructive madness created by Cameron Todd Willingham.” Willingham transcript, *supra* note 109, vol. XIII, at 45.

concluded: “On first reading, a contemporary fire origin and cause analyst might well wonder how anyone could make so many critical errors in interpreting the evidence.”<sup>127</sup> Nevertheless, a stay was denied, and Willingham was put to death. Subsequent evaluations agreed that the trial evidence was junk science. For example, five independent experts prepared a forty-three page report, finding that “each and every one of the indicators relied upon have since been scientifically proven to be invalid.”<sup>128</sup>

In May 2006, the Innocence Project petitioned the Texas Forensic Science Commission (TFSC) to review the arson testimony in Willingham’s and Ernest Ray Willis’ cases.<sup>129</sup> The TFSC is not authorized to determine guilt or innocence. Instead, the Innocence Project argued that the State Fire Marshall Office should have reinvestigated arson cases in which its experts testified after NFPA 921 was published in 1992 — a full twelve years before Willingham’s execution.<sup>130</sup> TFSC retained its own independent consultant, Dr. Craig Beyler, another nationally-recognized expert, to review the arson evidence. His fifty-one page report dissected the expert testimony, concluding:

The investigations of the Willis and Willingham fires did not comport with either the modern standard of care expressed by NFPA 921, or the standard of care expressed by fire investigation texts and papers in the period 1980-1992. The investigators had poor understandings of fire science and failed to acknowledge or apply the contemporaneous understanding of the limitations of fire indicators. Their methodologies did not comport with the scientific method or the process of elimination. A finding of arson could not be sustained based upon the standard of care expressed by NFPA, or the standard of care expressed by fire investigation texts and papers in the period 1980-1992.<sup>131</sup>

Once Beyler’s report became public, a political firestorm erupted, and the

---

<sup>127</sup> Report of Dr. Gerald Hurst, In re Cameron Todd Willingham, Trial Court No. 24, 4670(B), District Court, 366th Jud. Dist., Navarro County, Tex., Feb. 13, 2004.

<sup>128</sup> CARPENTER ET AL., *supra* note 119.

<sup>129</sup> The expert evidence in both cases was comparable, but Willis was lucky. His death penalty conviction was overturned on procedural grounds, and the prosecutor subsequently refused to reindict him after Dr. Hurst wrote the same type of critical report in Willis’s case that he had written in Willingham’s. Willis, who had spent seventeen years on death row, was subsequently exonerated on actual innocence grounds. See Mary Alice Robbins, *New-York Based Innocence Project Attacks Texas Arson Convictions*, 22 TEX. LAWYER, May 8, 2006.

<sup>130</sup> See Letter from Innocence Project to Texas Forensic Science Comm’n (Aug. 20, 2010).

<sup>131</sup> CRAIG L. BEYLER, ANALYSIS OF THE FIRE INVESTIGATION METHODS AND PROCEDURES USED IN THE CRIMINAL ARSON CASES AGAINST EARNEST RAY WILLIS AND CAMERON TODD WILLINGHAM, Aug. 17, 2009, at 51 (emphasis added).



governor, who was in the midst of a reelection battle, abruptly replaced commission members three days before a meeting scheduled to consider the Beyler report.<sup>132</sup> The newly-appointed chair, a prosecutor, promptly cancelled the meeting,<sup>133</sup> raising the specter of a cover-up.<sup>134</sup> Next, the Attorney General issued an opinion ruling that the TFSC lacked jurisdiction over cases decided before its creation.<sup>135</sup>

The TFSC eventually produced a report—one that did not directly deal with the Willingham and Willis cases. Nevertheless, the report's recommendations and statements indicated that the Willingham arson investigation was seriously flawed. Its first recommendation was "that fire investigators adhere to the standards of NFPA 921."<sup>136</sup> In addition, the report reviewed a number of arson indicators that were used in the Willingham and Willis cases. Citing Vasquez's testimony, the report undermined his opinions concerning (1) V-patterns as an indicator of origin, (2) pour patterns, (3) low/deep burning, (4) multiple separate points of origin, (5) spalling, (6) burn intensity, and (7) crazed glass.<sup>137</sup> It also observed that "testimony, such as Vasquez's response to a question regarding Willingham's state of mind, is an example of the type of testimony that experts should avoid as falling outside of their field of

---

<sup>132</sup> See Christy Hoppe, *Perry Defends Removing 3: He Says He's Following Protocol, but Critics Believe He's Derailing Arson Inquiry*, DALLAS MORNING NEWS, Oct. 2, 2009, at 3A; Mary Alice Robbins, *Fired Up: Changes Sought for Texas Forensic Science Commission at Center of Heated Controversy*, 25 TEX. LAWYER, Nov. 9, 2009 ("[Former Commissioner] Levy says he believes 'things went south' for the commission after [former Chair] Bassett released Beyler's report to the public in August 'as he was required by law to do.'"). The meeting was scheduled for October 2, 2009.

<sup>133</sup> Hoppe, *supra* note 132 (noting that the new chair was "known as one of the toughest law-and order prosecutors in the state").

<sup>134</sup> See Jennifer Emily, *Texas Forensic Science Commission Refuses to End Inquiry into Willingham Arson Case*, DALLAS MORNING NEWS, Sept. 18, 2010 ("Perry's replacements were seen by some as a political maneuver intended to change the outcome of the commission's decision."); Christy Hoppe, *Perry Ousts Officials Before Arson Hearing: He's Assailed as New Chair Delays Session on Flawed Case that Led to Execution*, DALLAS MORNING NEWS, Oct. 1, 2009, at 1A; David Mann, *Fire and Innocence*, TEX. OBSERVER, Dec. 3, 2009 ("Then in late September, Perry booted three members off of the Texas Forensic Science Commission, which was investigating the Willingham and Willis cases, just three days before a crucial hearing on scientists' findings. Perry's new appointees promptly canceled the hearing and have yet to reschedule it. Even conservative commentators cried cover-up, suggesting that Perry, in a tough battle for re-election, was trying to subvert an investigation that might prove he oversaw the execution of an innocent man.").

<sup>135</sup> Letter from Greg Abbott, Texas Attorney General, to Texas Forensic Science Comm'n (Aug. 2011).

<sup>136</sup> REPORT OF THE TEXAS FORENSIC SCIENCE COMM'N, WILLINGHAM/WILLIS INVESTIGATION 39 (April 15, 2011).

<sup>137</sup> *Id.* at 21-28.

expertise.”<sup>138</sup> The report even encouraged lawyers to “aggressively pursue admissibility hearings in arson cases.”<sup>139</sup>

Despite the opinions of all the independent experts, the State Fire Marshall vigorously defended its investigation. In a breathtaking letter, the office asserted that “[i]n reviewing the documents and standards in place then *and now*, we stand by the original investigator’s report and conclusions.”<sup>140</sup> This left the TFSC incredulous.<sup>141</sup>

### 3. *Han Tak Lee Case*

Unfortunately, Willingham’s case was not an outlier. In the 1989 trial of Han Tak Lee,<sup>142</sup> the expert also relied on the old “myths” to declare the fire incendiary: (1) greater intensity and heat, (2) burn patterns, (3) alligatoring, (4) melted metal in bed frames, and (5) crazed glass.<sup>143</sup> In addition, the investigation was “hobbled by an incomplete and inaccurate understanding” of flashover. After serving twenty-five years, Lee was released from prison in 2015.<sup>144</sup>

### 3. *National Fire Protection Association Guidelines*

After the publication of NFPA 921 in 1992, the kind of testimony presented in the Willingham and Lee cases should have vanished from the courtroom. But arson investigators balked. According to one expert, “[t]he initial

---

<sup>138</sup> *Id.* at 36.

<sup>139</sup> *Id.* at 48.

<sup>140</sup> Letter from Paul Maldonado, State Fire Marshal, to Leigh Tomlin, Commission Coordinator, Texas Forensic Science Comm’n (Aug. 20, 2010) (emphasis added).

<sup>141</sup> REPORT OF THE TEXAS FORENSIC SCIENCE COMM’N, *supra* note 136, at 16 (“This appears to be an untenable position in light of advances in fire science. The fires in these cases occurred two decades ago; there are few circumstances in which an investigation could not be improved with the benefit of twenty years of controlled scientific experiment and practical experience.”).

<sup>142</sup> Lee petitioned for a writ of habeas corpus in 2010, based in part on “inaccurate and unreliable evidence.” *Han Tak Lee v. Tennis*, No. 4:CV-08-1972, 2010 WL 3812160, at \*2 (M.D. Pa. Sept. 22, 2010). Although the district court denied Lee’s petition, the Third Circuit reversed. *Han Tak Lee v. Glunt*, 667 F.3d 397, 407-08 (3d Cir. 2012) (“If Lee’s expert’s independent analysis of the fire scene evidence — applying principles from new developments in fire science — shows that the fire expert testimony at Lee’s trial was fundamentally unreliable, then Lee will be entitled to federal habeas relief on his due process claim.”).

<sup>143</sup> *Han Tak Lee v. Tennis*, Civil No. 4:08-CV-1972, 2014 WL 3894306 (M.D. Pa. June 13, 2014) (magistrate report), *aff’d sub. nom.*, *Han Tak Lee v. Houtzdale*, 798 F.3d 159 (3d Cir. 2015).

<sup>144</sup> Mark Hansen, *Badly Burned: Long-Held Beliefs About Arson Science Have Been Debunked After Decades of Misuse and Scores of Wrongful Convictions*, 101 A.B.A. J. 37 (Dec. 2015).



response to NFPA 921 in the fire investigation community was overwhelmingly negative.<sup>145</sup> *Babick v. Berghuis*<sup>146</sup> is illustrative. In that case, Andrew Babick was convicted of arson-murder for a 1995 house fire and was sentenced to two terms of life imprisonment without the possibility of parole. He later sought habeas relief, claiming ineffective assistance of counsel and prosecutorial misconduct. In 2010, the Sixth Circuit rejected these claims.

However, in dissent, Judge Merritt chastised the defense attorney for not contesting the arson evidence in “this strange junk science case.”<sup>147</sup> One prosecution expert testified that: (1) char marks on the porch were evidence of an accelerant, (2) a “line of demarcation” in a burn pattern on a carpet was “suspicious” because “it should not have burned the carpeting on these jagged edges,” and (3) the burns were “not normal” and were “unnatural.”<sup>148</sup> Another prosecution expert stated that “low burning” and other “unnatural” patterns indicated the presence of an accelerant. Both experts “testified — in direct contrast to the NFPA guide — that they were so confident in their reading of burn patterns that the absence of any laboratory confirmation of accelerant had no effect on their testimony.”<sup>149</sup>

#### 4. Dog-sniff Evidence

More alarming, in Judge Merritt’s view, was dog-sniff evidence. The NFPA guide provides: “Research has shown that canines have been alerted to pyrolysis products that are not produced by an ignitable liquid” and a positive canine alert without laboratory confirmation “should not be considered validated.”<sup>150</sup> The lab tests had not detected accelerants in the house debris. Yet, a dog handler testified that “his dog, Samantha, was ‘1000 times’ more effective at detecting fire starters or liquid accelerants than a laboratory test on burnt material.”<sup>151</sup> In short, the “jury was misled into trusting Samantha over the arson forensic lab.”<sup>152</sup>

---

<sup>145</sup> *Id.*

<sup>146</sup> 620 F.3d 571 (6th Cir. 2010). See generally Marc Price Wolf, *Habeas Relief from Bad Science: Does Federal Habeas Corpus Provide Relief for Prisoners Possibly Convicted on Misunderstood Fire Science?*, 10 MINN. J.L., SCI. TECH. 213 (2009).

<sup>147</sup> 620 F.3d at 580.

<sup>148</sup> *Id.* at 581 (quoting transcript).

<sup>149</sup> *Id.*

<sup>150</sup> NFPA 921, *supra* note 105, § 16.5.4.7 (describing the role of canine investigations as “assisting with the location and collection of samples” for laboratory testing).

<sup>151</sup> *Babick*, 620 F.3d at 580.

<sup>152</sup> *Id.* See also *United States v. Myers*, No. 3:10-00039, 2010 WL 2723196 (S.D.W. Va. July 8, 2010) (granting motion in limine to prohibit expert testimony of a canine handler because the alert had not been confirmed by lab testing, conflicted with the Fire Guide,

A more recent arson-dog case involved James Hebshie, who was convicted of arson and mail fraud in 2006. A federal district court granted his habeas petition based on ineffective assistance of counsel grounds.<sup>153</sup> In the court's view, had a *Daubert* hearing been requested on the canine evidence, there was a "'reasonable probability' that the Court would have excluded the canine testimony or severely limited it."<sup>154</sup> Without a challenge from the defense, the dog handler testified that his dog (Billy) was 97% accurate.<sup>155</sup> Indeed, the handler testified to "an almost mystical account of Billy's powers and her unique olfactory capabilities."<sup>156</sup> The court explained: "[The handler] went on and on about what he understood about Billy, as if his relationship with Billy somehow enhanced the reliability and probative value of the results — that she was unique, that he could 'read her face,' that he was with her 365 days a year, that he knew her personality, 'the way her eyes shifted,' the ways her ear shifted, etc."<sup>157</sup>

The handler focused on one area as the origin of the fire and testified that the dog had not alerted anywhere else on the premises. However, the handler had limited the dog's access to that one area. In addition, a dog's failure to alert has no evidential value: "[T]he scientific literature cast doubt on the significance of the dog's *failure* to alert (false negatives) and even raised concerns about canine 'proficiency' testing, concerns counsel never raised."<sup>158</sup> Indeed, the term "accelerant-detection" dog was misleading because the dog is trained to alert to many common materials that are not accelerants; the site of the fire was a convenience store which sold lighter fluid and lighters.

### 5. Post-Daubert Cases

The courts' response to bogus arson evidence is mixed.<sup>159</sup> It is not hard to find cases citing discredited arson indicators after *Daubert*, such as pour patterns

---

and did not meet the *Daubert* standards).

<sup>153</sup> United States v. Hebshie, 754 F. Supp. 2d 89 (D. Mass. 2010).

<sup>154</sup> *Id.* at 124.

<sup>155</sup> See Michael E. Kurtz et al., *Effect of Background Interference on Accelerant Detection by Canines*, 41 J. FORENSIC SCI. 868 (1996) (discussing the varying levels of reliability in accelerant detection depending on the substance in question and the canine handler); see also *Farm Bureau Mut. Ins. Co. v. Foote*, 14 S.W.3d 512, 518 (Ark. 2000) (affirming the trial court's exclusion of a canine handler who sought to testify about "the alleged superior ability of his canine partner, Benjamin, to detect the presence of accelerants after a fire . . . [that he could] discriminate between different types of chemicals," and that he had an accuracy rate of "100%").

<sup>156</sup> *Hebshie*, 754 F. Supp. 2d at 93.

<sup>157</sup> *Id.* at 119.

<sup>158</sup> *Id.* at 94.

<sup>159</sup> I GIANNELLI ET AL., *supra* note 15, at 1102-03 ("Many appellate courts continue to routinely accept investigators' testimony about experientially base generalizations.") (citing cases).



or puddle configurations,<sup>160</sup> melted bedsprings,<sup>161</sup> concrete spalling,<sup>162</sup> fire load,<sup>163</sup> and “fast and hot” burn.<sup>164</sup> Decided in 1998, *Michigan Millers Mut. Ins. Corp. v. Benfield*<sup>165</sup> is considered the “first serious challenge to the ‘old school’ of fire investigators.”<sup>166</sup> In that case, the Eleventh Circuit ruled that arson testimony “is subject to *Daubert*’s inquiry regarding the reliability of such testimony”<sup>167</sup> and cited NFPA 921.<sup>168</sup> Yet, a 2011 article on the subject began with the passage: “Fire researchers have shattered dozens of arson myths in recent years. So why do

---

<sup>160</sup> See, e.g., *State v. Allen*, No. 22835, 2009 WL 2096295 ¶ 114 (Ohio Ct. App. 2009) (investigator testifying to “an irregular burn pattern on the floor which through all my experience and training it appears to be an irregular pour patterns [sic], an ignitable liquid pour pattern”); *State v. Wolf*, 891 N.E.2d 358, 360 (Ohio Ct. App. 2008) (A firefighter testified “that he observed ‘pour patterns’ located on the floor throughout the mobile home; that the pour patterns are burnt marks that look like puddles that result from ignitable liquids . . . being poured out of containers . . . .”); *Colburn v. State*, 990 So. 2d 206, 209-10 (Miss. Ct. App. 2008) (“This pour pattern, [the fire investigator] explained, was indicative of flammable liquid being poured in the area . . . . On cross-examination [the fire investigator] did admit that the State Crime Laboratory was unable to identify ignitable liquids in the three debris samples taken from the pour pattern area.”); *State v. Henderson*, 125 P.3d 1132, 1137 (Mont. 2005) (finding that the trial court “did not err in allowing [a firefighter] to identify in the photographs and diagrams the pour patterns he had observed at the scene”).

<sup>161</sup> *Simon v. State*, 633 So. 2d 407, 409 (Miss. 1993), *vacated*, *Simon v. Mississippi*, 513 U.S. 956 (1994).

<sup>162</sup> See, e.g., *State v. Amodio*, 915 A.2d 569, 576 (N.J. Super. Ct. App. Div. 2007) (“They washed the floor and observed areas of spalling in the concrete underneath the door. This was an indication that a flammable liquid had been employed in that area.”); *McCord v. Gulf Guar. Life Ins. Co.*, 698 So. 2d 89, 95 (Miss. 1997) (“The arson investigator . . . testified that he found five different areas of spalling and concluded arson to be the cause of the fire.”).

<sup>163</sup> See, e.g., *Wise v. State*, 719 N.E.2d 1192, 1200 (Ind. 1999) (A fire investigator testified that a fire was intentionally set based on several factors, including that “the fire burned too fast for its fuel load.”); *Carter v. State*, 516 S.E.2d 556, 560 (Ga. Ct. App. 1999) (The fire investigator “deduced there must have been an accelerant or some kind of extra fuel load.”).

<sup>164</sup> See, e.g., *People v. Klait*, No. 06-000399-FH, 2010 WL 2076956, at \*5 (Mich. Ct. App. May 25, 2010) (“[T]hey both testified that they believed, based on the fast and hot nature of the fire, that it was set intentionally.”); *State v. Walters*, 813 P.2d 857, 858 (Idaho 1990) (A fire investigator testified that “it was a hot, fast fire as opposed to a small or as opposed to a slow, smoldering fire, yes, the evidence suggests to me that it was deliberately set.”); *State v. Cutlip*, No. 99-I-149, 2001 WL 687493, at \*2 (Ohio Ct. App. 2001) (A fire department lieutenant testified to a list of factors including that “the fire was fast and hot” and “that such observations are typical of a fire started by someone pouring an accelerant and lighting it.”).

<sup>165</sup> 140 F.3d 915, 920 (11th Cir. 1998).

<sup>166</sup> John J. Lentini, *The Evolution of Fire Investigations and Its Impact on Arson Cases*, 27 CRIM. JUST. 12, 14 (Spring 2012).

<sup>167</sup> *Benfield*, 140 F.3d at 920.

<sup>168</sup> See also *Fireman’s Fund Ins. Co. v. Canon U.S.A., Inc.*, 394 F.3d 1054, 1058 (8th Cir. 2005) (holding district court’s exclusion of expert arson evidence proper where experts failed to compare hypothesis to evidence from scene in violation of NFPA 921); *Ind. Ins. Co. v. Gen. Elec. Co.*, 326 F. Supp. 2d 844, 850-51 (N.D. Ohio 2004) (holding that cause-and-origin expert’s failure to properly collect evidence violated NFPA 921).

American courts still lag behind?”<sup>169</sup> And a 2013 survey of 586 public sector fire investigators found that some myths endure: “Nearly 40 percent did not know that crazed glass is caused by rapid cooling, not rapid heating. Twenty-three percent think puddle-shaped burns indicate the use of an accelerant. Eight percent still believe that alligator blistering implies that a fire burned fast and hot.”<sup>170</sup>

\* \* \*

The Texas Forensic Science Commission’s report did more than the courts to curb flawed arson testimony. And it took the execution of an innocent man to trigger that report.<sup>171</sup> In addition, the resistance to change is all-too-familiar: Rules based on science “were slow to take hold, as veteran investigators clung to what now are considered disproven theories. In some police and fire departments, investigators were openly hostile to the updated science.”<sup>172</sup>

#### D. Comparative Bullet Lead Analysis

For over thirty years, FBI experts testified about Comparative Bullet Lead Analysis (CBLA), a technique that was first used in the investigation into President Kennedy’s assassination.<sup>173</sup> CBLA compares trace chemicals found in bullets at crime scenes with ammunition found in the possession of a suspect. This technique was used when firearms (“ballistics”) identification could not be employed. FBI experts used various analytical techniques (first, neutron activation analysis (NAA), and then inductively coupled plasma-atomic emission spectrometry (ICP-AES)) to determine the concentrations of seven elements — arsenic, antimony, tin, copper, bismuth, silver, and cadmium — in the bullet lead alloy of both the crime-scene and suspect’s bullets. Statistical tests were then used to compare the elements in each bullet and determine whether the fragments and suspect’s bullets were “analytically indistinguishable” for each of the elemental concentration means. Exactly what the phrase “analytically indistinguishable” meant was the central issue — i.e., did such a finding mean that the bullet fragments came from a small or large universe? Obviously, the probative value of the test results would differ if only a hundred bullets had the same chemical composition as opposed to several million bullets.

---

<sup>169</sup> Douglas Starr, *Up in Smoke*, DISCOVER 36, 37 (Nov. 2011).

<sup>170</sup> Hansen, *supra* note 144, at 42-43.

<sup>171</sup> For a fuller discussion of the case, see Paul C. Giannelli, *Junk Science and the Execution of an Innocent Man*, 7 NYU J. LAW & LIBERTY 221 (2013).

<sup>172</sup> See Steve Mills, *Convicted Murderer Hopes Latest Fire Science Proves Innocence*, Chicago Trib., May 18, 2015.

<sup>173</sup> See generally Erik Randich & Patrick M. Grant, *Proper Assessment of the JFK Assassination Bullet Lead Evidence from Metallurgical and Statistical Perspectives*, 51 J. FORENSIC SCI. 717 (2006) (discussing the original analysis of the bullet fragments).



The published cases revealed disparate and often inconsistent interpretive conclusions provided by FBI experts. In some, experts testified only that two exhibits were “analytically indistinguishable.”<sup>174</sup> In other cases, examiners concluded that samples *could have* come from the same “source” or “batch.”<sup>175</sup> In still others, they stated that the samples *came* from the same source.<sup>176</sup> The testimony in numerous cases went much further and referred to a “box” of ammunition (usually 50 loaded cartridges, sometimes 20). For example, two specimens:

- (1) Could have come from the same box,<sup>177</sup>
- (2) Could have come from the same box or a box manufactured on the same day,<sup>178</sup>
- (3) Were consistent with their having come from the same box of ammunition,<sup>179</sup>
- (4) Probably came from the same box,<sup>180</sup> or
- (5) Must have come from the same box or from another box that would have been made by the same company on the same day.<sup>181</sup>

Several other statements that differ appear in the published opinions. An early case reported that the specimens “had come from the same batch of ammunition: they had been made by the same manufacturer on the same day and at the *same hour*.”<sup>182</sup> One case reports the expert’s conclusion with a statistic.<sup>183</sup> In another case, the expert used the expressions “such a finding is rare”<sup>184</sup> and “a

---

<sup>174</sup> See *Wilkerson v. State*, 776 A.2d 685, 689 (Md. 2001).

<sup>175</sup> See *State v. Krummacher*, 523 P.2d 1009, 1012-13 (Or. 1974).

<sup>176</sup> See *United States v. Davis*, 103 F.3d 660, 673-74 (8th Cir. 1996); *People v. Lane*, 628 N.E.2d 682, 689-90 (Ill. App. Ct. 1993).

<sup>177</sup> See *State v. Strain*, 885 P.2d 810, 817 (Utah Ct. App. 1994); *State v. Jones*, 425 N.E.2d 128, 131 (Ind. 1981).

<sup>178</sup> See *State v. Grube*, 883 P.2d 1069, 1078 (Idaho 1994); *People v. Johnson*, 499 N.E.2d 1355, 1366 (Ill. 1986).

<sup>179</sup> See *State v. Reynolds*, 297 S.E.2d 532, 534 (N.C. 1982).

<sup>180</sup> See *Bryan v. State*, 935 P.2d 338, 360 (Okla. Crim. App. 1997).

<sup>181</sup> See *United States v. Davis*, 103 F.3d 660, 666-67 (8th Cir. 1996) (“An expert testified that such a finding is rare and that the bullets must have come from the same box or from another box that would have been made by the same company on the same day.”); *Commonwealth v. Daye*, 587 N.E.2d 194, 207 (Mass. 1992); *State v. King*, 546 S.E.2d 575, 584 (N.C. 2001) (The expert “opined that, based on her lead analysis, the bullets she examined either came from the same box of cartridges or came from different boxes of the same caliber, manufactured at the same time.”).

<sup>182</sup> *Brown v. State*, 601 P.2d 221, 224 (Alaska 1979) (emphasis added).

<sup>183</sup> *State v. Earhart*, 823 S.W.2d 607, 614 (Tex. Crim. App. 1991).

<sup>184</sup> *United States v. Davis*, 103 F.3d 660, 666 (8th Cir. 1996).

very rare finding.”<sup>185</sup> In still another case, the expert “opined that the same company produced the bullets at the same time, using the same lead source. Based upon Department of Justice records, she opined that an overseas company called PMC produced the bullets around 1982.”<sup>186</sup>

### 1. *NAS Bullet Lead Report (2004)*

The technique was not seriously challenged until a retired FBI examiner, William Tobin, began questioning the procedure in scientific and legal journals<sup>187</sup> and in court testimony as well.<sup>188</sup> As a result, the FBI asked the National Academy of Sciences to review the technique. The 2004 NAS report undercut the FBI testimony: “The available data do not support any statement that a crime bullet came from a particular box of ammunition. In particular, references to ‘boxes’ of ammunition in any form should be avoided as misleading under Federal Rule of Evidence 403.”<sup>189</sup> Perhaps the most disturbing case is *State v. Earhart*,<sup>190</sup> a capital murder case in which the CBLA evidence apparently played a

---

<sup>185</sup> *Id.* at 667.

<sup>186</sup> *People v. Villarta*, No. H021354, 2002 WL 66887 (Cal. Ct. App. Jan. 17, 2002) (murder). In later years, the testimony became more limited. A 2002 FBI publication states the conclusion as follows: “Therefore, they likely originated from the same manufacturer’s source (melt) of lead.” Charles A. Peters, *The Basis for Compositional Bullet Lead Comparisons*, 4 FORENSIC SCI. COMM. No. 3, at 5 (July 2002) (emphasis added). Testimony to the same effect has also been proffered. Transcript of Record at 6, *Commonwealth v. Wilcox*, No. 00CR2727 (Ky. Cir. Ct. Jefferson County Feb. 28, 2002) (trial testimony of Charles Peters, FBI examiner): “Well, bullets that are analytically indistinguishable likely come from the same molten lead sources of lead, uh, as opposed to bullets that have different composition come from different, uh, melts of lead.”

<sup>187</sup> See Edward J. Imwinkelried & William A. Tobin, *Comparative Bullet Lead Analysis (CBLA) Evidence: Valid Inference or Ipse Dixit?*, 28 OKLA. CITY U. L. REV. 43 (2003); Erik Randich et al., *A Metallurgical Review of the Interpretation of Bullet Lead Compositional Analysis*, 127 FORENSIC SCI. INT’L 174 (2002) (Tobin was a coauthor); William A. Tobin & Wayne Duerfeldt, *How Probative is Comparative Bullet Lead Analysis?*, 17 CRIM. JUSTICE 26 (Fall 2002).

<sup>188</sup> *E.g.*, *Ragland v. Commonwealth*, 191 S.W.3d 569, 577 (Ky. 2006); *Clemons v. State*, 896 A.2d 1059, 1070, 1068 (Md. 2006); *State v. Behn*, 868 A.2d 329, 339-40 (N.J. Super. Ct. 2005) (Tobin’s affidavit submitted).

<sup>189</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCES, FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE 6 (2004). The author served on the NAS Committee.

<sup>190</sup> 823 S.W.2d 607, 614 (Tex. Crim. App. 1991) (“[The expert] concluded that the likelihood that two .22 caliber bullets came from the same batch, based on *all* the .22 bullets made in one year, is approximately .000025 percent, ‘give or take a zero.’ He subsequently acknowledged, however, that the numbers which he used to reach the .000025 percent statistic failed to take into account that there are different types of .22 caliber bullets made each year . . . .22, .22 long, and .22 long rifle. [The expert] ultimately testified that there could be several hundred thousand bullets per batch, but with some variation in the elemental composition within the



significant role.<sup>191</sup> The transcript contains the following expert testimony: “We can - from my 21 years experience of doing bullet lead analysis and doing research on boxes of ammunition down though the years I can determine if bullets came from the same box of ammunition . . .”<sup>192</sup> However, the NAS report found that the amount of bullets that can be produced from a melt “can range from the equivalent of as few as 12,000 to as many as 35 million 40grain, .22 caliber long rifle bullets.”<sup>193</sup> Earhart was subsequently executed.<sup>194</sup>

## 2. Post-Report Developments

Much of the FBI testimony rested on a database, which the Bureau had built up over the course of many years. Although the NAS committee frequently asked for this data during its year-long investigation, the FBI did not turn over the data until it was too late to include an analysis of the information in its report.<sup>195</sup> The two statisticians who served on the NAS committee later wrote that their subsequent inspection of the data “identified several peculiarities.”<sup>196</sup> First, the database was incomplete. The FBI claimed to have a “complete data file” of some 71,000+ measurements but only 64,869 were turned over. Moreover, only

---

batch.”) (emphasis added).

<sup>191</sup> See *Earhart v. Johnson*, 132 F.3d 1062, 1067 (5th Cir. 1998) (denying habeas relief, the court noted: “Given the significant role the bullet evidence played in the prosecution’s case, we shall therefore assume Earhart could have made a sufficient threshold showing that he was entitled to a defense expert under Texas law.”).

<sup>192</sup> Transcript of Record at 5248-49, *State v. Earhart*, No. 4064, Dist. Ct. Lee County, 21st Judicial Dist., Texas (testimony of John Riley). See also *id.* at 5258 (“Well, bullets that are - that have analytically indistinguishable compositions or compositions that are generally similar typically are found within the same box of ammunition and that is the case that we have here. Now, bullets that are the same composition can also be found in other boxes of ammunition, but it’s most likely those boxes would have been manufactured at the same place on or about the same date.”). But see testimony of Charles Peters, FBI examiner, *Commonwealth v. Wilcox*, Kentucky, Feb. 28, 2002 (*Dambert* hearing: “We have never testified, to my knowledge, that that bullet came from that box. We’d never say that. All we are testifying is that bullet, or that victim fragment or something, the bullet, either came from that box or the many boxes that were produced at the same time.” Transcript at 1-2) (emphasis added).

<sup>193</sup> NATIONAL RESEARCH COUNCIL, *supra* note 189, at 6.

<sup>194</sup> See Death Penalty Information Center, Searchable Database of Executions, <http://www.deathpenaltyinfo.org/views-executions> (search for “Earhart” under “Find Person” search box) (last visited Feb. 15, 2017).

<sup>195</sup> See Cliff H. Spiegelman & Karen Kafadar, *Data Integrity and the Scientific Method: The Case for Bullet Lead Data as Forensic Evidence*, 19:2 CHANCE 16, 22 (2006) (“During the open sessions of the committee meetings, the FBI claimed to have a ‘complete data file’ of some 71,000+ measurements. Following repeated requests from the Committee, the FBI submitted at its last meeting a CD-ROM that contained two data files with a combined total of 64,869 bullet (not 71,000+) measurement records. . . . This data set could not be analyzed in time for the release of the report. . . .”).

<sup>196</sup> *Id.*



measurements made by ICP-AES were included; a different analytical method, NAA, had been used before 1997. Both techniques measured the same elements, and therefore the results from either technique would have been suitable for comparison. Further, the numbering system for the bullets was “highly inconsistent and rather unexpected,” suggesting that some bullet measurements had been deleted.<sup>197</sup> Additionally, “a rough investigation of the measurement error indicated many measurement errors that exceeded the FBI’s claimed analytical precision of 2-5%.”<sup>198</sup> Finally, “only 15% of the 1079 cases listed in these two files had measurements from [National Institute of Standards and Technology] . . . making it impossible to determine the frequency of ‘matches’” in some cases.<sup>199</sup> Accordingly, the “missing data and the inconsistent precisions” undermined the Bureau’s public claims.<sup>200</sup> These authors were puzzled by the FBI’s failure to disclose data: “The scientific method is important for science generally; forensic science is no exception. . . . [T]he evidence in this paper suggest that, at least for [CBLA], forensic science failed in the requirement to share the material, methods and data to reach conclusions with the scientific community.”<sup>201</sup>

The FBI’s response to the NAS report was also disconcerting. The Bureau quickly put out a press release, obscuring the report’s findings.<sup>202</sup> The release highlighted the committee’s conclusion that the FBI was using appropriate instrumentation and suitable elements for comparison. Yet, these aspects of CBLA were never seriously questioned. Rather, the interpretation of the data was disputed. Only one sentence in the press release addressed this critical issue: “Recommendations by the [NAS] include suggestions to improve the statistical analysis, quality control procedures, as well as expert testimony.”<sup>203</sup> The news media read the report quite differently – e.g., “Study Shoots Holes in Bullet

---

<sup>197</sup> *Id.* (“[T]he numbering system of the bullets was highly inconsistent and rather unexpected, e.g., the bullets from a suspect in a particular case might be numbered Q13A, Q13B, Q13C, Q14A, Q14B, Q14C, . . . , leading one to wonder what happened to bullets Q01, Q02, . . . , Q12.”). Other illustrations of incomplete data were noted: “[W]hile most of the bullets indicated 3 measurements, about 30 bullets had six or more measurements.” *Id.* “[O]nly about 50% of the bullets in this data set were identified as having come from one of the four major bullet manufacturers in the United States (Cascade Cartridge, Inc.; Federal; Remington; Winchester); the ‘complete data file’ of 71,000 bullets may yield a higher proportion of bullets from these four manufacturers.” *Id.*

<sup>198</sup> *Id.*

<sup>199</sup> *Id.*

<sup>200</sup> *Id.*

<sup>201</sup> *Id.* at 22-23.

<sup>202</sup> Department of Justice, FBI News Release, Feb. 10, 2004.

<sup>203</sup> *Id.*

Analysis By FBI,<sup>204</sup> “Report Finds Flaws,”<sup>205</sup> “Panel Questions FBI Bullet Analysis,”<sup>206</sup> and “Report Questions the Reliability of an F.B.I. Ballistics Test.”<sup>207</sup>

The Bureau also included the following passage in the press release: “The basis of bullet lead compositional analysis is supported by approximately 50 peer-reviewed articles found in scientific publications beginning in the early 1970’s. Published research and validation studies have continued to demonstrate the usefulness of the measurements of trace elements within bullet lead.”<sup>208</sup> In contrast, the NAS report pointed out that there were “very few peer-reviewed articles on homogeneity and the rate of false positive matches” and “outside reviews have only recently been published.”<sup>209</sup>

Over a year later, the FBI discontinued CBLA testing<sup>210</sup> and issued another (and similar) press release. Once again, the release minimized the problems, citing the following reason for its decision: “While the FBI Laboratory still firmly supports the scientific foundation of bullet lead analysis, given the costs of maintaining the equipment, the resources necessary to do the examination, and its relative probative value, the FBI Laboratory has decided that it will no longer conduct this exam.”<sup>211</sup> Nevertheless, a month earlier, Dwight Adams, the laboratory director, had written a private memorandum to the FBI Director specifying different reasons for abandoning the technique, including the following comments: (1) “We cannot afford to be misleading to a jury” and (2) “We plan to discourage prosecutors from using our previous results in future cases.”<sup>212</sup> Neither concern was reflected in the press release.

In the wake of the NAS report, several state courts excluded CBLA

---

<sup>204</sup> Maurice Possley, *Study Shoots Holes in Bullet Analysis By FBI*, CHICAGO TRIB., Feb. 11, 2004, at 14.

<sup>205</sup> Charles Pillar, *Report Finds Flaws in FBI Bullet Analysis; Changes are Proposed for the Technique Often Cited in Expert Testimony in Criminal Trials*, L.A. TIMES, Feb. 11, 2004, at 12.

<sup>206</sup> Randolph E. Schmid, *Panel Questions FBI Bullet Analysis*, ASSOCIATED PRESS, Feb. 10, 2004.

<sup>207</sup> See also Eric Lichtblau, *Report Questions the Reliability of an F.B.I. Ballistics Test*, N.Y. TIMES, Feb. 11, 2004, at 22.

<sup>208</sup> FBI News Release, *supra* note 202.

<sup>209</sup> NATIONAL RESEARCH COUNCIL, *supra* note 189, at 100.

<sup>210</sup> Eric Lichtblau, *F.B.I. Abandons Disputed Test for Bullets From Crime Scene*, N.Y. TIMES, Sept. 2, 2005, at A12.

<sup>211</sup> Department of Justice, FBI News Release, Sept. 1, 2005.

<sup>212</sup> John Solomon, *FBI’s Forensic Test Full of Holes*, WASH. POST, Nov. 18, 2007, at A1.



evidence.<sup>23</sup> Surprisingly, the FBI supplied affidavits in several cases supporting prosecutors' efforts to sustain convictions based on the technique. In one affidavit, the FBI cited the NAS report but failed to mention that the report had faulted the Bureau's statistical methods. The chair of the NAS committee criticized the affidavit because it did "not discuss the statistical bullet-matching technique, which is key and probably the most significant scientific flaw found by the committee."<sup>24</sup> The affidavit was also misleading because it estimated that the maximum number of .22-caliber bullets in a batch of lead was 1.3 million, when the NAS committee found that the number could be as high as 35 million.<sup>25</sup>

On November 18, 2007, *60 Minutes* aired a segment on CBLA.<sup>26</sup> In an interview, the FBI lab director, now retired, acknowledged that testimony about boxes was "misleading and inappropriate."<sup>27</sup> That broadcast, along with a *Washington Post* investigation, questioned the FBI's response to the NAS report. The main problem was that only the FBI had records of all the cases in which its experts had testified, and the Bureau had declined to disclose the names of those cases.<sup>28</sup> Instead, the Bureau relied on the NAS report, its own press releases, and pro forma letters sent to prosecution and defense organizations to notify defendants. This method of communication was grossly inadequate because the

---

<sup>23</sup> See *Ragland v. Commonwealth*, 191 S.W.3d 569, 580 (Ky. 2006) (noting that "[i]f the FBI Laboratory that produced the CBLA evidence now considers such evidence to be of insufficient reliability to justify continuing to produce it, a finding by the trial court that the evidence is both scientifically reliable and relevant would be clearly erroneous"); *Clemons v. State*, 896 A.2d 1059, 1070, 1078 (Md. 2006) ("CBLA is not admissible under the *Frye-Reed* standard because it is not generally accepted within the scientific community as valid and reliable."; "Based on the criticism of the processes and assumptions underlying CBLA, we determine that the trial court erred in admitting expert testimony based on CBLA because of the lack of general acceptance of the process in the scientific community."); *State v. Behn*, 868 A.2d 329, 331 (N.J. Super. Ct. 2005) (finding the technique was "based on erroneous scientific foundations").

*But see Commonwealth v. Fisher*, 870 A.2d 864, 871 (Pa. 2005) ("The CBLA evidence, at best, established a possible connection between Appellant and the bullets recovered from the victim's body."). See also *United States v. Davis*, 406 F.3d 505, 509 (8th Cir. 2005) ("Davis's trial counsel cannot be said to be ineffective for failing to challenge the FBI's methodology on a basis that was not advanced by the scientific community at the time of trial.").

<sup>24</sup> Solomon, *supra* note 212 (quoting Ken MacFadden).

<sup>25</sup> *Id.*

<sup>26</sup> *60 Minutes: Evidence of Injustice* (CBS television broadcast, Nov. 18, 2007).

<sup>27</sup> *Id.*

<sup>28</sup> Solomon, *supra* note 212, at A1 ("Hundreds of defendants sitting in prisons nationwide have been convicted with the help of an FBI forensic tool that was discarded more than two years ago. But the FBI lab has yet to take steps to alert the affected defendants or courts, even as the window for appealing convictions is closing . . .").

letters neither highlighted the problem, nor its significance.<sup>219</sup> A few days after the *60 Minutes* expose, Senator Patrick Leahy, the Chairman of the Senate Judiciary Committee, sent a letter to the FBI Director noting that the Bureau's letters gave "the false impression that these discredited tests had continuing reliability."<sup>220</sup>

\* \* \*

Here, the flaws are many: Lack of foundational research, failure to make a database available to outside scientists, and ignoring the FBI's own protocols by presenting inconsistent and misleading testimony. Moreover, the reluctance to confess error and take timely corrective action violated basic scientific norms. After decades of use, a federal district court in 2003 excluded CBLA evidence under the *Daubert* standard<sup>221</sup> for the first time.

### III. MISLEADINGLY PRESENTED TECHNIQUES

#### A. Firearms & Toolmark Identifications

Firearms identifications, popularly known as "ballistics," is another long-established forensic discipline. It developed in the early part of the last century, and by the 1930s courts were admitting evidence based on this technique. Subsequent cases followed these precedents, admitting evidence of bullet, cartridge case, and shot shell identifications.<sup>222</sup> Toolmark comparison, a related discipline, was also accepted during this period.<sup>223</sup> At the time *Daubert* was decided, the FBI's position was clear: "Firearms identification is the Forensic Science discipline that identifies a bullet, cartridge case or other ammunition component as having been fired by a particular firearm *to the exclusion of all other firearms*."<sup>224</sup> Yet, the examination, by means of a comparison microscope, is subjective and without a meaningful standard.

---

<sup>219</sup> The Innocence Network and the National Association of Criminal Defense Lawyers formed a task force and worked with the FBI to contact defense attorneys and convicts. See Vesna Jaksic, *Faulty Bullet-Test Cases Finding Way to Court*, *NAFL L.J.*, Feb. 25, 2008 ("The task force is lining up pro bono commitments from several law firms to handle the cases.").

<sup>220</sup> John Solomon, *Leahy Pursues Forensic Test Answers; Attorney General Is Told to Prepare For Senate Inquiry*, *WASH. POST*, Nov. 22, 2007, at A2 (quoting Leahy also wrote: "The new revelations about bullet-lead analysis are just the latest examples of the Department's inadequate efforts to ensure that sound forensic testing is utilized to the maximum extent to find the guilty rather than merely obtain a conviction. Punishing the innocent is wrong and allows the guilty party to remain free." *Id.*

<sup>221</sup> *United States v. Mikos*, No. 02 CR 137, 2003 WL 22922197 (N.D. Ill. Dec. 9, 2003).

<sup>222</sup> I GIANNELLI ET AL., *supra* note 15, § 14.06.

<sup>223</sup> *Id.* at § 14.12.

<sup>224</sup> FBI HANDBOOK OF FORENSIC SCIENCE 57 (rev. ed. 1994) (emphasis added).



## 1. Post-Daubert Cases

The courts gave short shrift to the initial post-*Daubert* challenges to firearms and toolmark identifications.<sup>225</sup> In 2005, however, the legal landscape changed abruptly. In *United States v. Green*,<sup>226</sup> the district judge questioned the foundational basis of firearms identifications. The court wrote that the expert “declared that this match could be made ‘to the exclusion of every other firearm in the world.’ . . . . That conclusion, needless to say, is extraordinary, particularly given [his] data and methods.”<sup>227</sup> Moreover, the expert could not cite any reliable error rates and admitted that he relied mainly on his subjective judgment. In addition, “[t]here were no reference materials of any specificity, no national or even local database on which he relied. And although he relied on his past experience with these weapons, he had no notes or pictures memorializing his past observations.”<sup>228</sup> In the end, the court restricted the expert’s testimony; he could only explain the ways in which the casings were similar but not that they came from a specific weapon “to the exclusion of every other firearm in the world.” In the court’s view, that conclusion “stretches well beyond [the expert’s] data and methodology.”<sup>229</sup>

A few weeks later, a different district judge in *United States v. Monteiro*<sup>230</sup> found that the technique “is largely a subjective determination [and] based on experience and expertise.”<sup>231</sup> Importantly, the court also concluded that the theory on which the expert relied was “tautological.” The Association of Firearm and Toolmark Examiners (AFTE), the leading organization of examiners, proposed the theory.<sup>232</sup> Under this theory, the examiner may declare an identification if (1) there is “sufficient agreement” of marks between the crime scene and test bullets

---

<sup>225</sup> See, e.g., *United States v. Hicks*, 389 F.3d 514, 526 (5th Cir. 2004) (stating that “the matching of spent shell casings to the weapon that fired them has been a recognized method of ballistics testing in this circuit for decades”); *United States v. Foster*, 300 F. Supp. 2d 375, 377 n.1 (D. Md. 2004) (“Ballistics evidence has been accepted in criminal cases for many years. . . . In the years since *Daubert*, numerous cases have confirmed the reliability of ballistics identification.”); *United States v. Santiago*, 199 F. Supp. 2d 101, 111 (S.D.N.Y. 2002) (“The Court has not found a single case in this Circuit that would suggest that the entire field of ballistics identification is unreliable.”).

<sup>226</sup> 405 F. Supp. 2d 104 (D. Mass. 2005).

<sup>227</sup> *Id.* at 107.

<sup>228</sup> *Id.*

<sup>229</sup> *Id.* at 109.

<sup>230</sup> 407 F. Supp. 2d 351 (D. Mass. 2006).

<sup>231</sup> *Id.* at 355.

<sup>232</sup> See *Theory of Identification*, Association of Firearm and Toolmark Examiners, 30 AFTE J. 86 (1998).

and (2) there is “sufficient agreement” when the examiner says there is.<sup>233</sup> In short, the “sufficient agreement” threshold is “in the minds eye of the examiner and is based largely on training and experience.”<sup>234</sup> The court would not admit the evidence unless the expert could better document the examination.

Together, *Green* and *Monteiro* should have served as a shot across the bow. But they did not; courts continued to admit the same evidence as before.<sup>235</sup>

## 2. NAS Ballistic Imaging Report (2008)

In 2008, the National Academy of Sciences published a report on computer imaging of bullets.<sup>236</sup> Although firearms identification was not the primary focus of the investigation, a section of the report commented on the subject.<sup>237</sup> After surveying the literature on uniqueness, reproducibility, and permanence of individual characteristics, the report noted that “[m]ost of these studies are limited in scale and have been conducted by firearms examiners (and examiners in training) in state and local law enforcement laboratories as adjuncts to their regular casework.”<sup>238</sup> The report found that the “validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related

---

<sup>233</sup> See Itiel E. Dror, *How Can Francis Bacon Help Forensic Science? The Four Idols of Human*, 50 JURIMETRICS 93, 104 (2009) (“The potential problem here is the nonscientific nature of the identification criteria. If the comparison of toolmarks enables conclusions about common origin when the unique surface contours of two toolmarks are in ‘sufficient agreement,’ what is the scientific definition and measurement of what constitutes such ‘sufficient agreement’? It seems that it is more in the eye of the beholder than strict scientific measures because it is determined without specific quantification and criteria.”).

<sup>234</sup> *Monteiro*, 407 F. Supp. 2d at 370.

<sup>235</sup> See, e.g., *United States v. Williams*, 506 F.3d 151, 161-162 (2d Cir. 2007) (upholding admissibility of firearms identification evidence—bullets and cartridge casings); *United States v. Natson*, 469 F. Supp. 2d 1253, 1261 (M.D. Ga. 2007) (“According to his testimony, these toolmarks were sufficiently similar to allow him to identify Defendant’s gun as the gun that fired the cartridge found at the crime scene. He opined that he held this opinion to a 100% degree of certainty . . . . The Court also finds [expert’s] opinions reliable and based upon a scientifically valid methodology. Evidence was presented at the hearing that the toolmark testing methodology he employed has been tested, has been subjected to peer review, has an ascertainable error rate, and is generally accepted in the scientific community.”).

<sup>236</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCES, BALLISTIC IMAGING (2008).

<sup>237</sup> The committee was asked to assess the feasibility, accuracy, reliability, and technical capability of developing and using a national ballistic database as an aid to criminal investigations. It concluded: (1) “A national reference ballistic image database of all new and imported guns is not advisable at this time.” (2) The National Integrated Ballistics Information Network (NIBIN) “can and should be made more effective through operational and technological improvements.” *Id.*

<sup>238</sup> *Id.* at 70.

toolmarks has not yet been fully demonstrated.”<sup>239</sup> The report went on to caution:

Conclusions drawn in firearms identification should not be made to imply the presence of a firm statistical basis when none has been demonstrated. Specifically, . . . examiners tend to cast their assessments in bold absolutes, commonly asserting that a match can be made “to the exclusion of all other firearms in the world.” Such comments cloak an inherently subjective assessment of a match with an extreme probability statement that has no firm grounding and unrealistically implies an error rate of zero.<sup>240</sup>

Citing this report, the district court in *United States v. Glynn*<sup>241</sup> ruled that the expert would only be permitted to testify that it was “more likely than not” that recovered bullets and cartridge cases came from a particular weapon.<sup>242</sup> The court also commented: “Based on the *Daubert* hearings . . . , the Court very quickly concluded that whatever else ballistics identification analysis could be called, it could not fairly be called ‘science,’”<sup>243</sup> further noting that “[t]he problem is compounded by the tendency of ballistics experts . . . to make assertions that their matches are certain beyond all doubt, that the error rate of their methodology is ‘zero,’ and other such pretensions.”<sup>244</sup>

### 3. *NAS Forensic Science Report (2009)*

As noted earlier, NAS issued its forensic report the following year in 2009. That report summarized the state of the research as follows:

Because not enough is known about the variabilities among individual tools and guns, we are not able to specify how many points of similarity are necessary for a given level of confidence in the result. Sufficient studies have not been done to understand the reliability and repeatability of the methods. . . . Individual patterns from manufacture or from wear might, in some cases, be distinctive enough to suggest one particular source, but additional studies should be performed to make the process of

---

<sup>239</sup> *Id.* at 81. The report also stated: “Additional general research on the uniqueness and reproducibility of firearms-related toolmarks would have to be done if the basic premises of firearms identification are to be put on a more solid scientific footing.” *Id.* at 82.

<sup>240</sup> *Id.* at 82.

<sup>241</sup> 578 F. Supp. 2d 567 (S.D.N.Y. 2008).

<sup>242</sup> *Id.* at 575.

<sup>243</sup> *Id.* at 570.

<sup>244</sup> *Id.* at 574.



individualization more precise and repeatable.<sup>245</sup>

In a different passage, the report — citing firearm and toolmark identifications observed that “[i]n]uch forensic evidence . . . is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline.”<sup>246</sup>

AFTE rejected these findings out of hand, arguing that NAS “ignore[d] extensive research supporting the scientific underpinnings of the identification of firearm and toolmark evidence.”<sup>247</sup> The court in *United States v. Otero*<sup>248</sup> accepted the AFTE’s position, citing studies which it was ill-equipped to evaluate.<sup>249</sup> A subsequent review of the oft-cited studies by two scientists concluded:

Exaggerated and unfounded implications relating to rates of error inferred from even the best of existing experiments in the field of firearms/toolmarks, generally self-described as ‘validation studies’, typically result from statistical, metallurgical and/or psychological (cognitive) deficiencies in the design and conduct of the experiments, and frequently lead to unjustified inferential extrapolation to universal assumption for the practice domain.<sup>250</sup>

Other courts took an important, but still limited, step of restricting examiner testimony by precluding the expert from making gross overstatements such as declaring a match to the exclusion, either practical or absolute, of all other weapons.<sup>251</sup> Similarly, some courts forbade experts from testifying that they hold

---

<sup>245</sup> NAS FORENSIC REPORT, *supra* note 21, at 154.

<sup>246</sup> *Id.* at 107-08.

<sup>247</sup> *The Response of the Association of Firearms and Tool Mark Examiners to the February 2009 National Academy of Science Report “Strengthening the Forensic Science in the United States: A Path Forward.”* 41 AFTE J. 204, 206 (2009).

<sup>248</sup> 849 F. Supp. 2d 425, 437-38 (D.N.J. 2012) (“The Court’s analysis of the proposed testimony according to the *Daubert* factors leads it to conclude that [the] expert report and opinion are admissible under Rule 702.”).

<sup>249</sup> See *infra* notes 138-39 (PCAST report).

<sup>250</sup> Clifford H. Spiegelman & William A. Tobin, *Analysis of Experiments in Forensic Firearms/Toolmarks Practice Offered as Support for Low Rates of Practice Error and Claims of Inferential Certainty*, 13 LAW, PROB. & RISK 115, 115 (2013).

<sup>251</sup> See, e.g., *United States v. Asburn*, 88 F. Supp. 3d 239, 249 (E.D.N.Y. 2015) (“Nor can [the expert] testify that a match he identified is to ‘the exclusion of all other firearms in the world,’ or that there is a ‘practical impossibility’ that any other gun could have fired the recovered materials.”); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1180 (D.N.M. 2009) (“[The expert] also will not be allowed to testify that he can conclude that there is a match to the exclusion, either practical or absolute, of all other guns.”).



their opinions to a “reasonable degree of scientific certitude.”<sup>252</sup> That term has long been required by courts in many jurisdictions for the admission of expert testimony. Incredibly, the phrase has no scientific meaning and the claim of certainty is unsupported by empirical research. Thus, it is grossly misleading. Indeed, the National Commission on Forensic Science rejected it.<sup>253</sup> Still other courts went off on a quixotic tangent, substituting the phrase “reasonable degree of ballistic” certitude.<sup>254</sup> Changing “scientific certainty” to “ballistic certainty” merely underscores the courts’ scientific incompetence.

However, even these modest limitations were rejected by other courts.<sup>255</sup> For example, in *United States v. Casey*,<sup>256</sup> the district court declined “to follow sister courts who have limited expert testimony based upon the 2008 and 2009 NAS reports and, instead, remains faithful to the long-standing tradition of allowing the unfettered testimony of qualified ballistics experts.”<sup>257</sup>

---

<sup>252</sup> See, e.g., *Asburn*, 88 F. Supp. 3d at 249 (“[T]he court joins in precluding this expert witness from testifying that he is ‘certain’ or ‘100%’ sure of his conclusions that certain items match.”); *United States v. Willock*, 696 F. Supp. 2d 536, 549 (D. Md. 2010) (“[The expert] shall state his opinions and conclusions without any characterization as to the degree of certainty with which he holds them.”); *People v. Robinson*, 2 N.E.3d 383, 402 (Ill. App. Ct. 2013) (“[T]he judicial decisions uniformly conclude toolmark and firearms identification is generally accepted and admissible at trial. Accordingly, we conclude the trial court did not err in ruling the testimony in this case was admissible . . . , particularly where the trial judge barred the witnesses from testifying their opinions were ‘within a reasonable degree of scientific certainty.’”).

<sup>253</sup> Nat’l Comm’n on Forensic Sci., Department of Justice, Views Document on Use of the Term “Reasonable Scientific Certainty” (adopted at NCFS Meeting #9 March 22, 2016).

<sup>254</sup> *Taylor*, 663 F. Supp. 2d at 1180 (“He may only testify that, in his opinion, the bullet came from the suspect rifle to within a reasonable degree of certainty in the firearms examination field.”); *United States v. Cerna*, No. CR 08–0730 WHA, 2010 WL 3448528, at \* 4 (N.D. Cal. 2010) (“a reasonable degree of certainty in the ballistics field”); *Commonwealth v. Pytou Heang*, 942 N.E.2d 927, 945 (Mass. 2011) (stating that “the expert may offer that opinion to a ‘reasonable degree of ballistic certainty’”).

<sup>255</sup> See, e.g., *Fleming v. State*, 1 A.3d 572, 590 (Md. Ct. App. 2010) (“[N]otwithstanding the current debate on the issue, courts have consistently found the traditional method [of firearms identification] to be generally accepted within the scientific community, and to be reliable.”); *People v. Givens*, 912 N.Y.S.2d 855, 857 (Sup. Ct. 2010) (“This Court was unable to find any cases where firearms and toolmark identification was found to be unreliable or no longer scientifically acceptable.”).

<sup>256</sup> 928 F. Supp. 2d 397, 400 (D. Puerto Rico 2013).

<sup>257</sup> *Id.* at 400. See also *United States v. Sebbem*, No. 10 Cr. 87(SL.T), 2012 WL 5989813 (E.D.N.Y. Nov. 30, 2012); *State v. Langlois*, 2 N.E.3d 936, ¶ 41 (Ohio Ct. App. 2014) (“Our conclusion on this issue finds support in the decisions of other appellate districts in Ohio, notwithstanding the recent criticisms in scientific reports and the limitations some federal courts have imposed on the testimony of firearms experts. These decisions hold that the methodology of comparatively analyzing and testing bullets and shell cases recovered from crime scenes is reliable.”); *State v. Jones*, 303 P.3d 1084, ¶ 75 (Wash. Ct. App. 2013) (expert testimony comparing bunter marks on the base of shell casings found at the crime scene to shell casings found in Jones’s home admissible under *Frve* standard).

#### 4. White House PCAST Report (2016)

The 2016 White House PCAST report agreed with the NAS 2009 report's characterization of the scientific research on firearms and toolmarks identification: "We find that many of these earlier studies were inappropriately designed to assess foundational validity and estimate reliability. Indeed, there is internal evidence among the studies themselves indicating that many previous studies underestimated the false positive rate by at least 100-fold."<sup>258</sup> In addition, PCAST found only one of the post-2009 studies sufficiently rigorous. The Defense Department's Forensic Science Center commissioned the study, which was conducted by an independent testing lab (the Ames Laboratory, a Department of Energy national laboratory affiliated with Iowa State University). In this study, "[t]he false-positive rate was estimated at 1 in 66, with a confidence bound indicating that the rate could be as high as 1 in 46."<sup>259</sup> The study has not been published in a scientific journal. According to the PCAST report, more than one study is required and studies should be published in peer-reviewed scientific literature. Consequently, "the current evidence still falls short of the scientific criteria for foundational validity."<sup>260</sup>

The AFTE quickly retorted, expressing their "disappointment in the PCAST's choice to ignore the research that has been conducted" and claiming that "[d]ecades of validation and proficiency studies have demonstrated that firearm and toolmark identification is scientifically valid."<sup>261</sup> However, when PCAST later invited stakeholders to submit validation studies that it may have overlooked, no studies satisfying PCAST's criteria were offered.<sup>262</sup>

\* \* \*

The lessons here are familiar. For years, an entrenched forensic discipline vigorously guarded its turf by rejecting the conclusions of the outside scientific community.<sup>263</sup> It published a journal which was "peer-reviewed" by other

---

<sup>258</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 11.

<sup>259</sup> *Id.*

<sup>260</sup> *Id.*

<sup>261</sup> *Association of Firearm and Tool Mark Examiners Response to PCAST Report on Forensic Science*, 48 AFTE J. 195, 195 (2016).

<sup>262</sup> PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, AN ADDENDUM TO THE PCAST REPORT ON FORENSIC SCIENCE IN CRIMINAL COURTS, Jan. 6, 2017, at 7 ("Several respondents wrote to PCAST concerning firearms analysis. None cited additional appropriately designed black-box studies similar to the recent Ames Laboratory study.").

<sup>263</sup> See William A. Tobin et al., *Absence of Statistical and Scientific Ethos: The Common Denominator in Deficient Forensic Practices*, 3 STATISTICS & PUBLIC POLICY (Dec. 16, 2016) ("[P]ractitioners remain intractable even after years of critical scholarly papers, ad hoc committees of the National Academy of Sciences (NAS), position statements from the U.S.



members of its discipline. The journal, which is advertised as “the Scientific Journal” of AFTE, was not generally available until 2016. The discipline claimed to be a “science” but did not hold itself to the normative standards of science. The AFTE “Theory of Identification” is “clearly not a scientific theory, which the National Academy of Sciences has defined as ‘a comprehensive explanation of some aspect of nature that is supported by a vast body of evidence. . . .’ More importantly, the stated method is circular.”<sup>264</sup> Only recently, after two NAS reports, have some courts begun to limit misleading testimony. Many have not. Thus, the courts’ competence to deal with flawed research remains extant.<sup>265</sup>

In 2005, the district court in *Green* cautioned: “The more courts admit this type of toolmark evidence without requiring documentation, proficiency testing, or evidence of reliability, the more sloppy practices will endure; we should require more.”<sup>266</sup> Over a decade later, a concurring opinion in *Williams v. United States*<sup>267</sup> concluded: “As matters currently stand, a certainty statement regarding toolmark pattern matching has the same probative value as the vision of a psychic: it reflects nothing more than the individual’s foundationless faith in what he believes to be true.”<sup>268</sup> In short, there is a “lost decade” during which the discipline summarily dismissed criticisms when it should have lead the effort for more rigorous research.

## B. Fingerprint Examinations

Before DNA analysis, fingerprint identification was the gold standard in forensics.<sup>269</sup> Like many other forensic disciplines, it gained judicial acceptance decades before *Daubert* was decided. *People v. Jennings*,<sup>270</sup> the first reported fingerprint case, was decided in 1911. In 1984, the FBI pronounced the technique “infallible” in its official publication, which also referred to the technique as a “science.”<sup>271</sup> Nevertheless, it is a subjective technique without an objective

---

Department of Justice . . .”).

<sup>264</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 6.

<sup>265</sup> Tobin et al., *supra* note 263 (“the purported ‘validation studies’ typically proffered to courts are seriously flawed [and] have no external validity”).

<sup>266</sup> *Green*, 405 F. Supp. 2d at 109.

<sup>267</sup> 130 A.3d 343 (D.C. 2016).

<sup>268</sup> *Id.* at 355 (concurring).

<sup>269</sup> See Joseph L. Peterson & Anna S. Leggett, *The Evolution of Forensic Science: Progress Amid the Pitfalls*, 36 STETSON L. REV. 621, 654 (2007) (“The scientific integrity and reliability of DNA testing have helped DNA replace fingerprinting and made DNA evidence the new ‘gold standard’ of forensic evidence.”).

<sup>270</sup> 96 N.E. 1077 (Ill. 1911). See generally: 1 CHANNELLI ET AL., *supra* note 15, ch. 18 (discussing the scientific and legal issues associated with fingerprint identification).

<sup>271</sup> FEDERAL BUREAU OF INVESTIGATION, THE SCIENCE OF FINGERPRINTS IV (1984).

standard and typically involves partial prints with inevitable distortions.

### 1. *Post-Daubert Cases*

After *Daubert*, challenges to fingerprint comparison testimony were decidedly unsuccessful.<sup>272</sup> One infamous case, *United States v. Havvard*,<sup>273</sup> illustrates the judiciary's lack of rigor in applying *Daubert*. Not only did the district court uphold the admissibility of fingerprint testimony, it described the technique as “the very archetype of reliable expert testimony under [the *Daubert/Kumho*] standards.”<sup>274</sup> According to the court, latent print identification had been “tested” for nearly 100 years in adversarial proceedings with the highest possible stakes — liberty and sometimes life. Yet, *Daubert* required scientific, not “adversarial,” testing.<sup>275</sup> Next, in citing “peer review,” the court noted that a second fingerprint examiner also compared the prints: “In fact, peer review is the standard operating procedure among latent print examiners.”<sup>276</sup> This statement reveals a fundamental misunderstanding of “peer review” as used in *Daubert*. In that case, peer review meant refereed scientific journals in which validation research is published. An amici brief submitted in *Daubert* by the *New England Journal of Medicine* and other scientific publications explained that peer review’s “role is to promote the publication of well-conceived articles so that the most important review, the consideration of the reported results by the scientific community, may occur after publication.”<sup>277</sup>

Moreover, the court accepted the prosecution expert’s astounding claim that the “error rate for the method is zero.”<sup>278</sup> Experts argued that, while individual examiners may make mistakes, the method itself is perfect. However, the dichotomy between “methodological” and “human” error rates in this context

---

<sup>272</sup> See, e.g., *United States v. Collins*, 340 F.3d 672, 682 (8th Cir. 2003) (“Fingerprint evidence and analysis is generally accepted.”); *United States v. Hernandez*, 299 F.3d 984, 991 (8th Cir. 2002); *United States v. Martinez-Cintrón*, 136 F. Supp. 2d 17, 20 (D. Puerto Rico 2001).

<sup>273</sup> 117 F. Supp. 2d 848 (S.D. Ind. 2000), *aff’d*, 260 F.3d 597 (7th Cir. 2001).

<sup>274</sup> *Id.* at 855.

<sup>275</sup> See Sandy L. Zabell, *Fingerprint Evidence*, 13 J.L. & POL’Y 143, 170 (2005) (The “argument that no latent print has ever been found to match the rolled print of a different person is . . . misleading because no systematic search for such pairs on the entire databank of millions of fingerprints has ever been performed.”).

<sup>276</sup> *Havvard*, 117 F. Supp. 2d at 854.

<sup>277</sup> Brief of the *New England Journal of Medicine*, *Journal of the American Medical Association*, and *Annals of Internal Medicine* as Amici Curiae in Support of Respondent, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993), at 3.

<sup>278</sup> *Havvard*, 117 F. Supp. 2d at 854.



is “practically meaningless”<sup>279</sup> because the examiner is the method.<sup>280</sup> Finally, the court turned *Daubert* on its head by requiring the defendant to prove the evidence was unreliable, a distortion that would be employed in later cases.<sup>281</sup>

Then, *United States v. Llera Plaza*<sup>282</sup> “sent shock waves through the community of fingerprint analysts.”<sup>283</sup> In that 2002 case, Judge Pollak ruled that fingerprint experts would not be permitted to testify that two sets of prints “matched” — that is, a positive identification to the exclusion of all other persons. This was apparently the first time in over 90 years that such a decision had been rendered.<sup>284</sup> On rehearing, however, Judge Pollak reversed himself,<sup>285</sup> and later cases continued to uphold the admissibility of fingerprint evidence.<sup>286</sup> Nevertheless, the case captured the attention of the media with news reports,<sup>287</sup>

---

<sup>279</sup> Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 BROOK. L. REV. 13, 60 (2001). Professor Mnookin goes on to provide this analogy: “The same argument could be made of eyewitness testimony, a notoriously unreliable form of evidence. People are all distinct from one another in observable ways; therefore the theoretical error rate of eyewitness identification is zero, though in practice observers may frequently makes errors.” *Id.* See also Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985, 1040 (2005) (stating that “in fingerprint practice the concept is vacuous”).

<sup>280</sup> See Zabell, *supra* note 275, at 172 (“But, given its unavoidable subjective component, in latent print examination people *are* the process.”).

<sup>281</sup> See Michael J. Saks, *The Legal and Scientific Evaluation of Forensic Science (Especially Fingerprint Expert Testimony)*, 33 SETON HALL L. REV. 1167, 1173-76 (2003) (discussing the reversal of the burden of persuasion as one of several judicial responses employed to avoid confronting the lack of empirical testing).

<sup>282</sup> 179 F. Supp. 2d 492 (E.D. Pa.), *vacated, mot. granted on recons.*, 188 F. Supp. 2d 549 (E.D. Pa. 2002).

<sup>283</sup> See D.H. Kaye, *The Nonscience of Fingerprinting: United States v. Llera-Plaza*, 21 QUINNIPIAC L. REV. 1073, 1073 (2003).

<sup>284</sup> As Professor Mnookin has noted, however, “fingerprints were accepted as an evidentiary tool without a great deal of scrutiny or skepticism.” Mnookin, *supra* note 279, at 17. She elaborated: “Even if no two people had identical sets of fingerprints, this did not establish that no two people could have a single identical print, much less an identical part of a print. These are necessarily matters of probability, but neither the court in *Jennings* nor subsequent judges ever required that fingerprinting identification be placed on a secure statistical foundation.” *Id.* at 19.

<sup>285</sup> 188 F. Supp. 2d 549, 572 (E.D. Pa. 2002).

<sup>286</sup> See, e.g., *United States v. Abreu*, 406 F.3d 1304, 1307 (11th Cir. 2005) (“We agree with the decisions of our sister circuits and hold that the fingerprint evidence admitted in this case satisfied *Daubert*.”); *United States v. Janis*, 387 F.3d 682, 690 (8th Cir. 2004) (finding fingerprint evidence reliable); *United States v. Mitchell*, 365 F.3d 215 (3d Cir. 2004); *United States v. Crisp*, 324 F.3d 261 (4th Cir. 2003); *United States v. Sullivan*, 246 F. Supp. 2d 700, 704 (E.D. Ky. 2003).

<sup>287</sup> E.g., Joann Loviglio, *Trial Judge Reaffirms Fingerprint Usability: Hearing Shows Him Science Involved*, SAN ANTONIO EXPRESS-NEWS, Mar. 14, 2002; Andy Newman, *Judge Who Ruled Out Matching Fingerprints Changes his Mind*, N.Y. TIMES, Mar. 14, 2002; Richard Willing, *Judge Challenges Fingerprint Identification*, USA TODAY, Jan. 30, 2002.

mainstream publications,<sup>288</sup> scientific journals,<sup>289</sup> and television shows giving it substantial coverage.<sup>290</sup> A spate of legal articles followed,<sup>291</sup> with some commentators believing that *Llera Plaza I* was more faithful to *Daubert* than *Llera Plaza II*.<sup>292</sup> In response, the FBI adopted a “circle the wagons” attitude, fiercely defending the technique. The head of the FBI fingerprint section told *60 Minutes* that the error rate was “zero”, examiners only testify to “hundred percent certainty,” and the FBI had won “forty-one out of forty-one” legal challenges to fingerprint evidence.<sup>293</sup>

The appellate opinion most faithful to *Daubert* appeared in *United States v. Crisp*<sup>294</sup> — unfortunately in dissent. The majority opinion upheld the admissibility of fingerprint evidence by shifting the burden of proof to the defendant and by grandfathering the technique.<sup>295</sup> In dissent, Judge Michael conscientiously applied the *Daubert* factors. First, he noted that the “government did not offer any record of testing on the reliability of fingerprint identification. . . . [T]here have not been any studies to establish how likely it is that partial prints taken from a crime scene will be a match for only one set of fingerprints in the world.”<sup>296</sup> Second, as for peer review, “[a]gain, the government offered no evidence on this factor at trial. Fingerprint examiners, . . . have their own professional publications. . . . But unlike typical scientific journals, the fingerprint publications do not run articles that include or prompt critique or reanalysis by other scientists. Indeed, few of the articles address the principles of

---

<sup>288</sup> E.g., Michael Specter, *Do Fingerprints Lie? The Gold Standard of Forensic Science is Now Being Challenged*, 78 THE NEW YORKER 96 (May 27, 2002) (discussing case including interview with judge).

<sup>289</sup> See David L. Faigman, *Is Science Different for Lawyers?*, 297 SCIENCE 339, 340 (2002).

<sup>290</sup> *60 Minutes: Fingerprints* (CBS television broadcast Jan. 5, 2003).

<sup>291</sup> See, e.g., Simon A. Cole, *Grandfathering Evidence: Fingerprint Admissibility Rulings from Jennings to Llera Plaza and Back Again*, 41 AM. CRIM. L. REV. 1189 (2004); Robert Epstein, *Fingerprints Meet Daubert: The Myth of Fingerprint “Science” Is Revealed*, 75 S. CAL. L. REV. 605 (2002); Kristin Romandetti, Note, *Recognizing and Responding to a Problem with the Admissibility of Fingerprint Evidence Under Daubert*, 45 JURIMETRICS 41 (2004).

<sup>292</sup> E.g., Jennifer L. Mnookin, *Fingerprints: Not a Gold Standard*, 20 ISSUES IN SCI. & TECH. 47 (2003) (“Judge Pollak’s first opinion [restricting latent fingerprint individualization testimony] was the better one.”); Recent Case, 115 HARV. L. REV. 2349, 2352 (2002) (“Fingerprint expert testimony does not survive application of the *Daubert* factors . . .”).

<sup>293</sup> *60 Minutes: Fingerprints* (CBS television broadcast Jan. 5, 2003).

<sup>294</sup> 324 F.3d 261 (4th Cir. 2003).

<sup>295</sup> *Id.* at 269 (“Put simply, Crisp has provided us no reason today to believe that this general acceptance of the principles underlying fingerprint identification has, for decades, been misplaced. Accordingly, the district court was well within its discretion in accepting at face value the consensus of the expert and judicial communities that the fingerprint identification technique is reliable.”).

<sup>296</sup> *Id.* at 273-74 (Michael, J., dissenting).



fingerprint analysis and identification at all . . . .”<sup>297</sup> Third, “an error rate must be demonstrated by reliable scientific studies, not by assumption.”<sup>298</sup> Fourth, “the government did not establish that there are objective standards in the fingerprint examination field to guide examiners in making their comparisons.”<sup>299</sup> Fifth, while acknowledging general acceptance in the fingerprint community, the judge remarked that “[n]othing in the record in this case shows that the fingerprint examination community has challenged itself sufficiently or has been challenged in any real sense by outside scientists.”<sup>300</sup>

## 2. Madrid Train Bombing

*Llera Plaza* was soon eclipsed by a more sensational event—the FBI’s misidentification of Brandon Mayfield as the source of the crime scene prints in the terrorist train bombing in Madrid on March 11, 2004.<sup>301</sup> More than any other event, the Mayfield affair exposed the myth of fingerprint infallibility. This debacle resulted in investigations by the FBI<sup>302</sup> and the Inspector General (IG) of the Department of Justice.<sup>303</sup> One of the more troubling aspects of these reports dealt with the culture in the laboratory. The FBI internal investigation found that “[t]o disagree was not an expected response,”<sup>304</sup> and the IG reported that “FBI examiners did not attempt to determine the basis of the [Spanish National Police’s] doubts before reiterating that they were ‘absolutely confident’ in the identification on April 15, a full week before the FBI Laboratory met with the SNP.”<sup>305</sup>

---

<sup>297</sup> *Id.* at 274.

<sup>298</sup> *Id.* The judge added: “In a 1995 test conducted by a commercial testing service, less than half of the fingerprint examiners were able to identify correctly all of the matches and eliminate the non-matches. On a similar test in 1998, less than sixty percent of the examiners were able to make all identifications and eliminations. . . . An error rate that runs remarkably close to chance can hardly be viewed as acceptable under *Daubert*.” *Id.* at 275.

<sup>299</sup> *Id.* at 276.

<sup>300</sup> *Id.*

<sup>301</sup> See Sara Kershaw, *Spain and U.S. at Odds on Mistaken Terror Arrest*, N.Y. TIMES, Jun. 5, 2004, at A1 (Spanish authorities cleared Brandon Mayfield and matched the fingerprints to an Algerian national); Flynn McRoberts & Maurice Possley, *Report Blasts FBI Lab: Peer Pressure Led to False ID of Madrid Fingerprint*, CHI. TRIB., Nov. 14, 2004, at 1.

<sup>302</sup> See Robert B. Stacey, *A Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 54 J. FORENSIC IDENTIFICATION 707 (2004).

<sup>303</sup> See OFFICE OF THE INSPECTOR GENERAL, U.S. DEP’T OF JUSTICE, A REVIEW OF THE FBI’S HANDLING OF THE BRANDON MAYFIELD CASE, UNCLASSIFIED EXECUTIVE SUMMARY 7 (2006) (“Having found as many as 10 points of unusual similarity, the FBI examiners began to ‘find’ additional features in [the print] that were not really there, but rather were suggested to the examiners by features in the Mayfield prints.”).

<sup>304</sup> Stacey, *supra* note 302, at 713.

<sup>305</sup> OFFICE OF THE INSPECTOR GENERAL, *supra* note 303, at 10.

In addition to highlighting the lack of foundational research, these events raised a host of other issues, including (1) the role of cognitive bias in subjective techniques,<sup>306</sup> (2) the lack of well-defined standards,<sup>307</sup> (3) the failure to administer rigorous proficiency tests,<sup>308</sup> (4) the manipulation of research,<sup>309</sup> and (5) other instances of misidentifications.<sup>310</sup> The FBI did not undertake a serious review of fingerprints until it was compelled to address the issue due to the negative publicity surrounding the Mayfield misidentification. Even then, however, the

---

<sup>306</sup> See Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT'L 74 (2006) (reporting an experiment that showed fingerprint examiners changed their opinions when provided with irrelevant information); Elizabeth F. Loftus & Simon A. Cole, Letter, *Contaminated Evidence*, 304 SCIENCE 959 (May 14, 2004) (“[F]orensic scientists remain stubbornly unwilling to confront and control the problem of bias, insisting that it can be overcome through sheer force of will and good intentions.”); Stacey, *supra* note 302, at 713 (“confirmation bias”). See generally 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, 39 (2002).

<sup>307</sup> Examiners follow a procedure known as Analysis, Comparison, Evaluation, and Verification (ACE-V). See Zabell, *supra* note 275, at 178 (“ACE-V is an acronym, not a methodology. It is merely the common sense description of what anyone would do if they were examining a latent and a candidate source print.”).

<sup>308</sup> See Crisp, 324 F.3d at 274 (4th Cir. 2003) (Michael, J., dissenting) (“Proficiency testing is typically based on a study of prints that are far superior to those usually retrieved from a crime scene.”); *Ilera Plaza*, 188 F. Supp. 2d at 565 (noting that “the FBI examiners got very high proficiency grades, but the tests they took did not. . . . [O]n the present record I conclude that the proficiency tests are less demanding than they should be.”); Jennifer L. Mnookin, Editorial, *A Blow to the Credibility of Fingerprint Evidence*, BOSTON GLOBE, Feb. 2, 2004 (“There are no systematic proficiency tests to evaluate examiners’ skill. Those tests that exist are not routinely used and are substandard.”).

<sup>309</sup> See Donald Kennedy, Editorial, *Forensic Science: Oxymoron?*, 302 SCIENCE 1625 (2003) (discussing the cancellation of a National Academies project designed to examine various forensic science techniques, including fingerprinting, because the Departments of Justice and Defense insisted on a right of review that the Academy had refused to other grant sponsors); *United States v. Mitchell*, 365 F.3d 215, 238 (3d Cir. 2004) (“We are deeply discomfited by Mitchell’s contention—supported by Dr. Rau’s account of events, though contradicted by other witnesses—that a conspiracy within the Department of Justice intentionally delayed the release of the solicitation until after Mitchell’s jury reached a verdict. Dr. Rau’s story, if true, would be a damning indictment of the ethics of those involved.”). See generally Paul C. Giannelli, *Daubert and Forensic Science: The Pitfalls of Law Enforcement Control of Scientific Research*, 2011 U. ILLINOIS L. REV. 53 (discussing the manipulation of forensic science research, including fingerprint research).

<sup>310</sup> See Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985 (2005) (collecting 23 cases involving mistakes). The misidentification cases include some that involved (1) verification by one or more other examiners, (2) examiners certified by the International Association of Identification, (3) procedures using a sixteen-point standard, and (4) defense experts who corroborated misidentifications made by prosecution experts. *Id.* at 985; *Reasonable Doubt: Can We Trust Crime Labs?*, CNN PRESENTS, Jan. 9, 2005 (discussing the misidentification of Rieki Jackson, who spent two years in prison).



FBI still characterized the technique as “scientific.”<sup>311</sup>

The scientific community continued to note the lack of research,<sup>312</sup> — and the courts continued to ignore this fact.<sup>313</sup> Indeed, in *United States v. Baines*,<sup>314</sup> decided in 2009, the head of the FBI fingerprint section testified: “As to these ‘false positives’ . . . the FBI had ‘made, on average, about one erroneous identification every 11 years.’ The total number of identifications made has been about one million per year, . . . so that the known actual error rate was about one per eleven million identifications.”<sup>315</sup> Problematically, he merely assumed that all the other identifications were correct, thus disqualifying his analysis. Perhaps the most troubling aspect of this testimony was the lack of self-awareness for a person who claimed to be a scientist.<sup>316</sup>

### 3. NAS Forensic Science Report (2009)

Fingerprint examiners follow a procedure known as Analysis, Comparison, Evaluation, and Verification (ACE-V). The 2009 NAS report observed that since “the ACE-V method does not specify particular measurements or a standard test protocol, . . . examiners must make subjective assessments throughout.”<sup>317</sup> Thus, the ACE-V method is too “broadly stated” to “qualify as a

---

<sup>311</sup> See Bruce Budowle et al., *Review of the Scientific Basis for Friction Ridge Comparisons as a Means of Identification: Committee Findings and Recommendations*, 8 FORENSIC SCI. COMM. (Jan. 2006).

<sup>312</sup> See Donald Kennedy & Richard A. Merrill, *Assessing Forensic Science*, 20 ISSUES IN SCI. & TECH. 33 (Fall 2003) (“The increased use of DNA analysis, which has undergone extensive validation, has thrown into relief the less firmly credentialed status of other forensic science identification techniques (fingerprints, fiber analysis, hair analysis, ballistics, bite marks, and tool marks). These have not undergone the type of extensive testing and verification that is the hallmark of science elsewhere.”); Zabell, *supra* note 275, at 164 (“Although there is a substantial literature on the uniqueness of fingerprints, it is surprising how little true scientific support for the proposition exists.”).

<sup>313</sup> See, e.g., *United States v. Pena*, 586 F.3d 105, 110 (1st Cir. 2009) (“The district court did not abuse its discretion. Numerous courts have found expert testimony on fingerprint identification based on the ACE-V method to be sufficiently reliable under *Daubert*.”); *United States v. Abreu*, 406 F.3d 1304, 1307 (11th Cir. 2005) (“[T]he fingerprint evidence admitted in this case satisfied *Daubert*.”); *United States v. Janis*, 387 F.3d 682, 690 (8th Cir. 2004) (“We conclude the district court did not err in admitting the fingerprint expert’s testimony.”); *United States v. Mitchell*, 365 F.3d 215, 241 (3d Cir. 2004).

<sup>314</sup> 573 F.3d 979 (10th Cir. 2009).

<sup>315</sup> *Id.* at 984.

<sup>316</sup> See WHITE HOUSE PCAST REPORT, *supra* note 33, at 53 (“The fallacy is obvious: the expert simply assumed without evidence that every error in casework had come to light.”).

<sup>317</sup> NAS FORENSIC REPORT, *supra* note 21, at 139.

validated method for this type of analysis.”<sup>318</sup> The report added that “[t]he latent print community in the United States has eschewed numerical scores and corresponding thresholds” and consequently relies “on primarily subjective criteria” in making the ultimate attribution decision.<sup>319</sup> In making the decision, the examiner must draw on his or her personal experience to evaluate such factors as “inevitable variations in pressure,” but to date those factors have not been “characterized, quantified, or compared.”<sup>320</sup> In addition, the report gave short shift to the zero-error-rate argument, finding that “claims that these analyses have zero error rates are not scientifically plausible.”<sup>321</sup> In conclusion, the report outlined an agenda for the research it considered necessary “[t]o properly underpin the process of friction ridge identification.”<sup>322</sup>

Several studies were published after the NAS report.<sup>323</sup> The most important was a FBI study published in 2011,<sup>324</sup> which is discussed below.

#### 4. *White House PCAST Report (2016)*

According to the White House PCAST report, “latent fingerprint analysis is a foundationally valid subjective methodology”<sup>325</sup> and the FBI “significantly advanced the field” by conducting the black-box study. Nevertheless, the false positive rate

is substantial and is likely to be higher than expected by many jurors based on longstanding claims about the infallibility of fingerprint analysis. The false-positive rate could be as high as 1 error in 306 cases based on the FBI study and 1 error in 18 cases based on a study by another crime

---

<sup>318</sup> *Id.* at 142.

<sup>319</sup> *Id.* at 141.

<sup>320</sup> *Id.* at 144. Moreover, examiners lack population frequency data to quantify how rare or common a particular type of fingerprint characteristic is. *Id.* at 144.

<sup>321</sup> *Id.* at 142. *See also id.* at 143 (“Some in the latent print community argue that the method itself, if followed correctly ... has a zero error rate. Clearly, this assertion is unrealistic ... . The method, and the performance of those who use it, are inextricably linked, and both involve multiple sources of error (e.g., errors in executing the process steps, as well as errors in human judgment.”).

<sup>322</sup> *Id.* at 144.

<sup>323</sup> *See* WHITE HOUSE PCAST REPORT, *supra* note 33, at 91-95.

<sup>324</sup> B.T. Ulery et al., *Accuracy and Reliability of Forensic Latent Fingerprint Decisions*, 108 PROC. NAT’L ACAD. SCI. 7733 (2011) (“To attempt to ensure that the non-mated pairs were representative of the type of matches that might arise when police identify a suspect by searching fingerprint databases, the known prints were selected by searching the latent prints against the 58 million fingerprints in the [Automated Fingerprint Identification System] database and selecting one of the closest matching hits.”).

<sup>325</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 9-10.

laboratory. In reporting results of [a] latent-fingerprint examination, it is important to state the false-positive rates based on properly designed validation studies[.]<sup>326</sup>

Moreover, “testimony asserting any specific level of increased accuracy (beyond that measured in the studies) due to blind independent verification would be scientifically inappropriate, as speculation unsupported by empirical evidence.”<sup>327</sup>

#### 5. AAAS Fingerprint Report (2017)

In September 2017, the American Association for the Advancement of Science published an extensive report on fingerprint analysis.<sup>328</sup> An accompanying news release, summarized the reports findings: “Courtroom testimony and reports stating or even those implying that fingerprints collected from a crime scene belong to a single person are indefensible and lack scientific foundation . . .”<sup>329</sup>

The report reached a number of conclusions. First, claims that experts can identify the source of a latent print with 100% accuracy, are “clearly overstated and are now widely recognized as indefensible.”<sup>330</sup> Second, use of the term “identification” in reports and testimony even with qualifications “fail to deal forthrightly with the level of uncertainty that exists in latent print examination” and “cannot be justified scientifically.”<sup>331</sup> Third, because of public misconceptions experts “should acknowledge: (1) that the conclusions being reported are opinions rather than facts (as in all pattern-matching disciplines), (2) that it is not possible for a latent print examiner to determine that two friction ridge impressions originated from the same source to the exclusion of all others; and (3) that errors have occurred in studies of the accuracy of latent print examination.”<sup>332</sup>

The report went on to make several recommendations. Experts should “avoid statements that claim or imply that the pool of possible sources is limited to a single person. Terms like ‘match,’ ‘identification,’ ‘individualization’ and

---

<sup>326</sup> *Id.* at 10.

<sup>327</sup> *Id.* at 99.

<sup>328</sup> AAAS FINGERPRINT REPORT, *supra* note 34.

<sup>329</sup> Ann Q. Hoy *Fingerprint Source Identity Lacks Scientific Basis for Legal Certainty: More Research into Validity of Fingerprint Comparisons Needed*, *Forensic Report Says*, Sept. 15, 2017.

<sup>330</sup> AAAS FINGERPRINT REPORT, *supra* note 34, at 9.

<sup>331</sup> *Id.* at 10.

<sup>332</sup> *Id.* at 11.



their synonyms, imply more than the science can sustain.”<sup>333</sup> In addition, experts should “be prepared to discuss forthrightly the results of research studies that tested the accuracy of latent print examiners on realistic known-source samples.”<sup>334</sup>

\* \* \*

Despite the ruckus created by *Ilera Plaza* and the Mayfield fiasco, examiner testimony remained unchanged. Testimony such as “zero error rates,” “matches to the exclusion of all other fingerprints,” and “100 percent certainty” — which had been used for decades — continued, while the fingerprint community remain oblivious that such statements were scientifically implausible. As with firearms identification, there is a “lost decade” during which more research could have been conducted. As one judge noted in a 2003 dissent: “The government has had ten years to comply with *Daubert*. It should not be given a pass in this case.”<sup>335</sup> Those words were written fourteen years ago.

On a positive note, the Mayfield incident did trigger the FBI’s black box study, which was a significant achievement. Still, this study was released 100 years after the courts first admitted fingerprint evidence.<sup>336</sup> The White House PCAST report found it “distressing” that properly constructed validation studies had only been conducted recently and only one study had been published in a peer-reviewed journal.<sup>337</sup> *Daubert* had little effect.<sup>338</sup>

---

<sup>333</sup> *Id.* The report suggested that “examiners might say something like the following:

‘The latent print on Exhibit ## and the record fingerprint bearing the name XXXX have a great deal of corresponding ridge detail with no differences that would indicate they were made by different fingers. There is no way to determine how many other people might have a finger with a corresponding set of ridge features, but this degree of similarity is far greater than I have ever seen in non-matched comparisons.’

*Id.* at 11.

<sup>334</sup> *Id.*

<sup>335</sup> *Crisp*, 324 F.3d at 272 (Michael, J., dissenting).

<sup>336</sup> See *supra* text accompanying note 270.

<sup>337</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 95.

<sup>338</sup> Some courts did place limitations on the testimony. See, e.g., *Mitchell*, 365 F.3d at 245-46 (“Testimony at the *Daubert* hearing indicated that some latent fingerprint examiners insist that there is no error rate associated with their activities or that the examination process is irreducibly subjective. This would be out-of-place under Rule 702.”); *Commonwealth v. Gombora*, 933 N.E.2d 50, 61 n.22 (Mass. 2010) (“opinions expressing absolute certainty about, or the infallibility of, an ‘individualization’ of a print should be avoided”).

#### IV. FORENSIC SCIENCE RESEARCH

By now it is almost a truism that too many forensic disciplines are not grounded in science — and yet their adherents continue to claim the mantle of science. The NAS report emphasized the “notable dearth of peer-reviewed, published studies establishing the scientific bases and validity of many forensic methods.”<sup>339</sup> Indeed, the co-chair of the NAS committee, Judge Harry Edwards, later stated: “I think that the most important part of our Committee’s Report is its call for real science to support the forensic disciplines.”<sup>340</sup> Not surprisingly, the report triggered extensive commentary.<sup>341</sup> One article cataloged the numerous ways in which forensic science has failed to develop a research culture<sup>342</sup> and argued that the “core values” of a scientific culture “are empiricism, transparency, and an ongoing critical perspective.”<sup>343</sup> Another article documented the serious problems that have arisen when the law enforcement controls forensic research.<sup>344</sup>

##### A. National Commission on Forensic Science (2013-17)

To its credit, the DOJ, in partnership with the National Institute of Standards and Technology (NIST), established the National Commission on Forensic Science in 2013. The commission’s task was to enhance the practice and improve the reliability of forensic science.<sup>345</sup> Early on, the commission created a subcommittee on scientific inquiry and research, which undertook the task of reviewing bibliographies of foundational literature that had been compiled by various forensic disciplines.<sup>346</sup> The subcommittee quickly concluded that even a

---

<sup>339</sup> NAS FORENSIC REPORT, *supra* note 21, at 8.

<sup>340</sup> Honorable Harry T. Edwards, The National Academy of Sciences Report on Forensic Sciences: What it Means for the Bench and Bar, Address at Conference of Superior Court of the District of Columbia, Washington, D.C., May 6, 2010, at 7.

<sup>341</sup> See Paul C. Giannelli, *The 2009 NAS Report: A Literature Review*, 48 CRIM. L. BULL. 378 (2012) (listing numerous articles and conferences).

<sup>342</sup> Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725 (2011).

<sup>343</sup> *Id.* at 742.

<sup>344</sup> See Paul C. Giannelli, *Daubert and Forensic Science: The Pitfalls of Law Enforcement Control of Scientific Research*, 2011 U. ILL. L. REV. 53.

<sup>345</sup> National Commission on Forensic Science, U.S. Department of Justice, <https://www.justice.gov/nfcs> (last visited ).

<sup>346</sup> As a result of the NAS report, an Interagency Working Group — the Research Development Technology and Evaluation (RDT&E) of the National Science and Technology Council’s Subcommittee on Forensic Science was tasked with identifying foundational research forensic sciences. National Science and Technology Council Committee on Science Subcommittee on Forensic Science, May 2, 2014, Office of Science & Technology Policy. The RDT&E committee requested Scientific Working Groups (SWG) with addressing a series of discipline-specific questions. In response, literature compendiums were submitted to the RDT&E



“cursory review” of the bibliographies raised serious concerns. One basic problem involved the definition of foundational literature. According to the subcommittee, “[i]n some cases, it was unclear which literature citations are crucial to support the foundation of a particular forensic discipline.”<sup>347</sup> This finding led the subcommittee to define the term: foundational, scientific literature should consist of “original research, substantive reviews of the original research, clinical trial reports, or reports of consensus development conferences.”<sup>348</sup> Tellingly, the subcommittee felt compelled to add: “While other forms of dissemination of research and practice (e.g., oral and poster presentations at meetings, workshops, personal communications, editorials, dissertations, theses, and letters to editors) play an important role in science, the open, peer-reviewed literature is what endures and forms a foundation for further advancements.”<sup>349</sup>

The subcommittee’s second concern was that “some of the cited literature had not undergone a rigorous peer-review process.”<sup>350</sup> Peer review by other members of a forensic discipline is not sufficient.<sup>351</sup> Many of the reviewers are not scientists, and there is the problem with role bias. According to the subcommittee, foundational research should be subjected to “rigorous peer review with independent external reviewers to validate the accuracy . . . [and] overall consistency with scientific norms of practice”<sup>352</sup> and “published in a journal that is searchable using free, publicly available search engines.”<sup>353</sup> With few exceptions,

---

committee by several forensic working groups.

<sup>347</sup> Nat’l Comm’n on Forensic Sci., U.S. Department of Justice, Views Document on Scientific Literature in Support of Forensic Science and Practice (adopted at NCFSS Meeting #5 January 30, 2015) [hereinafter Views Document on Scientific Literature].

<sup>348</sup> *Id.*

<sup>349</sup> *Id.*

<sup>350</sup> *Id.*

<sup>351</sup> See *United States v. Crisp*, 324 F.3d 261, 274 (4th Cir. 2003) (Michael, J., dissenting) (“Fingerprint examiners, . . . have their own professional publications. . . . But unlike typical scientific journals, the fingerprint publications do not run articles that include or prompt critique or reanalysis by other scientists. Indeed, few of the articles address the principles of fingerprint analysis and identification at all . . .”). See also Zabell, *supra* note 275, at 164 (“Although there is a substantial literature on the uniqueness of fingerprints, it is surprising how little true scientific support for the proposition exists.”).

<sup>352</sup> Views Document on Scientific Literature, *supra* note 347 (“Published in a journal that maintains a clear and publicly available statement of purpose that encourages ethical conduct such as disclosure of potential conflicts of interest integral to the peer review process.”).

<sup>353</sup> *Id.* Other publication requirements include: (1) “Published in a journal or book that has an International Standard Number (ISSN for journals; ISBN for books) and recognized expert(s) as authors (for books) or on its Editorial Board (for journals).” (2) “Published in a journal that is indexed in databases that are available through academic libraries and other services (e.g. JSTOR, Web of Science, Academic Search Complete, and SciFinder Scholar).”



the disciplines considered above have not satisfied these requirements.<sup>354</sup>

Another recommendation, one on technical merit, provides: “All forensic science methodologies should be evaluated by an independent scientific body to characterize their capabilities and limitations in order to accurately and reliably answer a specific and clearly defined forensic question.”<sup>355</sup> Significantly, the commission recommended that the NIST be the independent scientific evaluator within the justice system.

## B. White House PCAST Report (2016)

Unlike the commission, which had a broad mandate, the White House PCAST report focused only on the validation issue. It took pains to explain the concept of validation, noting that forensic methods must be based on empirical studies and be “repeatable, reproducible, and accurate, at levels that have been measured and are appropriate to the intended application.”<sup>356</sup> The report

---

<sup>354</sup> Another commission document provided guidance for evaluating scientific literature. Nat’l Comm. on Forensic Sci., Department of Justice, Views Document on Identifying and Evaluating Literature that Supports the Basic Principles of a Forensic Science Method or Forensic Science Discipline (adopted at NCFS Meeting #9 – March 22, 2016). Including:

- Is the problem or hypothesis clearly stated?
- Is the scope of the article clearly stated as appropriate (article, case study, review, technical note, etc.)?
- Is the literature review current, thorough, and relevant to the problem being studied?
- Does this work fill a clear gap in the literature or is it confirmatory and/or incremental?
- Are the experimental procedures clear and complete such that the work could be easily reproduced?
- Are the experimental methods appropriate to the problem?
- Are the methods fully validated to the necessary level of rigor (fit for purpose)?
- Are the data analysis and statistical methodology appropriate for the problem, and explained clearly so it can be reproduced?
- Are the experimental results clearly and completely presented and discussed?
- Are omissions and limitations to the study discussed and explained?
- Are the results and conclusions reasonable and defensible based on the work and the supporting literature?
- Are the citations and references complete and accurate?
- Are the references original (primary) and not secondary?
- Are funding sources and other potential sources of conflict of interest clearly stated?

<sup>355</sup> Nat’l Comm. on Forensic Sci., Department of Justice, Views Document on Technical Merit Evaluation of Forensic Science Methods and Practices (Adopted at NCFS Meeting #10 - June 21, 2016). Recommendation: Recommendation on Technical Merit Evaluation of Forensic Science Methods and Practice (adopted at NCFS Meeting #11 - September 12, 2016).

<sup>356</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 4-5. Here, “repeatable” means an examiner reaches the same result when analyzing the same sample. “Reproducible” means that different examiners reach the same result when analyzing the same sample. The term “accurate” means that “an examiner obtains correct results both (1) for samples from the same source (true positives) and (2) for samples from different sources (true negatives).” Finally, “reliability” means

recognized that forensic methods may be either objective or subjective. Foundational validity for objective methods “can be established by studying [and] measuring the accuracy, reproducibility, and consistency of each of its individual steps.”<sup>357</sup> By definition, this approach is not possible with subjective techniques because they involve significant human judgment. Consequently, validity and reliability for these methods must be based on “black-box studies” (as if a “black box” is in the examiner’s head), in which numerous examiners make decisions on many independent tests in order to determine error rates.<sup>358</sup>

Importantly, the report also specified what does *not* qualify as validation: “experience, judgment, good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) cannot substitute for actual evidence of foundational validity and reliability.”<sup>359</sup> Moreover, expressions of *confidence* by individual examiners or a consensus among practitioners about the accuracy cannot substitute for “error rates estimated from relevant studies.” In sum, empirical evidence is the “sine qua non” for establishing foundational validity.<sup>360</sup>

PCAST also recommended that NIST conduct scientific evaluations of the validity of current and new forensic technologies: “To ensure the scientific judgments are unbiased and independent, such evaluations should be conducted by an agency which has no stake in the outcome.”<sup>361</sup>

In response, DOJ released a statement criticizing the report — on the day of its release. According to DOJ, the PCAST report “does not mention numerous published research studies which seem to meet PCAST’s criteria for appropriately designed studies providing support for foundational validity. That omission discredits the PCAST report as a thorough evaluation of scientific validity.”<sup>362</sup> PCAST, in turn, invited all stakeholders to identify validity studies that it might have overlooked. “DOJ ultimately concluded that it had no additional studies for

---

repeatability, reproducibility, and accuracy. *Id.* at 47.

<sup>357</sup> *Id.* at 5.

<sup>358</sup> *Id.* at 5-6.

<sup>359</sup> *Id.*

<sup>360</sup> *Id.*

<sup>361</sup> *Id.* The NAS report considered NIST before recommending an independent agency but rejected the idea because, at that time, NIST had limited ties to forensic science. NAS FORENSIC REPORT, *supra* note 21, at 17.

<sup>362</sup> Department of Justice, Comment Letter on PCAST’s Report to the President on Forensic Science in Federal Criminal Courts: Ensuring Scientific Validity of Pattern Comparison Methods (Sept. 20, 2016), <http://www.crime-scene-investigator.net/PDF/fbi-response-to-forensic-science-in-federal-criminal-courts-ensuring-scientific-validity-of-pattern-comparison-methods.pdf>.



PCAST to consider.”<sup>361</sup> Nor did the more than 400 papers submitted by twenty-six respondents cause PCAST to change its positions. The bottom line remained: “In science, empirical testing is the only way to establish the validity and degree of reliability of such an empirical method. Fortunately, empirical testing of empirical methods is feasible. There is no justification for accepting that a method is valid and reliable in the absence of appropriate empirical evidence.”<sup>364</sup> However, most prior studies use “closed-set design.” In these studies, “the correct source of each questioned sample is always present; studies using the closed-set design have underestimated the false-positive and inconclusive rates by more than 100-fold.”<sup>365</sup>

#### IV. INDEPENDENT SCIENTIFIC REVIEW

As discussed above, the courts have too often failed to fulfill their “gatekeeper”<sup>366</sup> function under *Daubert*. However, the *Daubert* Court also suggested that the adversary system would serve as a complementary safeguard, noting that “[v]igorous 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.”<sup>367</sup>

Yet, these “traditional” means have also proved inadequate. After the release of the NAS report, some commentators focused on defense counsel’s incompetence.<sup>368</sup> Moreover, a 2009 study of the cases of 137 convicts exonerated by DNA profiling revealed that “[d]efense counsel rarely made any objections to the invalid forensic science testimony in these trials and rarely effectively cross-examined forensic analysts who provided invalid science testimony.”<sup>369</sup> One commentator summed it up this way:

---

<sup>363</sup> PRESIDENT’S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, AN ADDENDUM TO THE PCAST REPORT ON FORENSIC SCIENCE IN CRIMINAL COURTS 3 (Jan. 6, 2017), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensics\\_addendum\\_finalv2.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf).

<sup>364</sup> *Id.* at 4.

<sup>365</sup> *Id.* at 7.

<sup>366</sup> *Daubert*, 509 U.S. at 597 (“a gatekeeping role for the judge”).

<sup>367</sup> *Id.* at 596 (citing *Rock v. Arkansas*, 483 U.S. 44, 61 (1987)).

<sup>368</sup> See Gertner, *supra* note 31, at 790 (“[T]he NAS Report’s concerns will not be fully met until advocacy changes.”); D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Path Forward Fraught with Pitfalls*, 2010 UTAH L. REV. 225, 242 (“Criminal defense lawyers . . . are supposed to be the people who recognize bogus expert claims, challenge them, move to get them excluded, and undermine those that survive exclusion by knowledgeable, thorough, and telling cross-examination. On the whole, they don’t do any of these things very well.”).

<sup>369</sup> Garrett & Neufeld, *supra* note 89, at 89.



Unlike the extremely well-litigated civil challenges, the criminal defendant's challenge is usually perfunctory. Even when the most vulnerable forensic sciences – hair microscopy, bite marks, and handwriting – are attacked, the courts routinely affirm admissibility citing earlier decisions rather than facts established at a hearing. Defense lawyers generally fail to build a challenge with appropriate witnesses and new data. Thus, even if inclined to mount a *Daubert* challenge, they lack the requisite knowledge and skills, as well as the funds, to succeed.<sup>370</sup>

Although defense bar bears some responsibility for *Daubert*'s failure, there are limits to what can be expected of overburdened and chronically underfunded public defenders when dealing with expert testimony. Better training for defense counsel (which is sorely needed) is not sufficient. Similarly, access to defense experts (also sorely needed) will not be adequate.<sup>371</sup> Defense experts can challenge prosecution experts' methods and opinions but do not have the funds to conduct foundational research, nor can they act as independent evaluators of foundational research on an ongoing basis.

In addition, prosecutors are ethically obliged to avoid the use of flawed forensic testimony.<sup>372</sup> Yet, the National District Attorneys Association recently asserted that bite mark evidence is a "reliable science"<sup>373</sup> — an untenable position.

The justice system is incapable of providing this expertise. An alternative paradigm is needed. An independent scientific review is required. NAS has published the most authoritative and independent reviews of forensic science. In addition to the forensic report, NAS has issued reports on sound spectrometry ("voiceprints"),<sup>374</sup> two DNA reports,<sup>375</sup> polygraph testing,<sup>376</sup> and bullet lead

---

<sup>370</sup> Neufeld, *supra* note 20, at S110.

<sup>371</sup> See Paul C. Giannelli, *Ake v. Oklahoma: The Right to Expert Assistance in a Post-Daubert, Post-DNA World*, 89 CORNELL L. REV. 1305 (2004) (discussing the legal disputes over the scope of the *Ake* – e.g., whether it applied to non capital cases and to non psychiatric experts).

<sup>372</sup> See Paul C. Giannelli & Kevin McMunigal, *Prosecutors, Ethics, and Expert Witnesses*, 76 FORDHAM L. REV. 1493 (2007).

<sup>373</sup> See *supra* text accompanying note 71.

<sup>374</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCE, ON THE THEORY AND PRACTICE OF VOICE IDENTIFICATION (1979).

<sup>375</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCE, THE EVALUATION OF FORENSIC DNA EVIDENCE (1996); NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCE, NATIONAL ACADEMY OF SCIENCES, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992).

<sup>376</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCES, THE POLYGRAPH AND LIE DETECTION (2003).

analysis.<sup>377</sup> But NAS is not a governmental entity, and its work depends on funding. The justice system needs scientific expertise on a continuing basis and thus institutionalized.

The National Commission's proposal, endorsed by PCAST, tasked NIST with the responsibility of evaluating forensic disciplines on an ongoing basis.<sup>378</sup> It should be adopted. NIST has the expertise and independence for this task and has been increasingly involved in forensic research. There would be a cost, but litigating validity issues across the country at *Daubert* and *Frye* hearings also has a cost. Moreover, there is a significant expense associated with rectifying the past mistakes that occurred with hair,<sup>379</sup> bullet lead,<sup>380</sup> DNA,<sup>381</sup> and arson cases.

Unfortunately, the current Attorney General did not renew the commission's charter in April 2017.<sup>382</sup> The independent scientists on the commission objected to this action, writing:

The Justice Department now proposes to improve forensic science by moving its oversight and development to an office within the department. This is precisely the opposite of what was recommended by the National Academy of Sciences report and the NCFS. It is a step backwards, because it reinforces the conditions that contributed to the current problems, namely, placing this discipline within the control of law enforcement and prosecutors. The Justice Department is home to many dedicated public servants including scientists whose passion for justice is

---

<sup>377</sup> NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCES, FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE (2004).

<sup>378</sup> In 2005, Peter Neufeld proposed an institute of forensic science. Neufeld, *supra* note 20, at S113.

<sup>379</sup> See David R. Cameron, *Forum: Review of FBI Lab Suggests Huge Number of Wrongful Convictions*, NEW HAVEN REGISTER, April 26, 2015 ("The FBI review has identified roughly 2,500 cases that fit those criteria. The review is still in its early stages; thus far, it has considered 268 trials involving 284 defendants. It has found that lab examiners gave flawed testimony regarding the comparison of hairs in 257 of the 268 trials — more than 95 percent. Almost all of the examiners over that period — 26 of 28 — presented flawed testimony.").

<sup>380</sup> See *supra* note 220.

<sup>381</sup> See Lauren Kirchner, *Traces of Crime: How New York's DNA Techniques Became Tainted*, N.Y. TIMES, Sept. 4, 2017 (explaining that two controversial techniques have been discontinued); Spencer S. Hsu, *FBI Notifies Crime Labs of Errors Used in DNA Match Calculations since 1999*, WASH. POST, May 29, 2015 ("The FBI has notified crime labs across the country that it has discovered errors in data used by forensic scientists in thousands of cases to calculate the chances that DNA found at a crime scene matches a particular person, several people familiar with the issue said.").

<sup>382</sup> See Spencer S. Hsu, *Sessions Orders Justice Dept. to End Forensic Science Commission, Suspend Review Policy*, WASH. POST, April 10, 2017.



unquestioned. However, DOJ is not a scientific body, and it is difficult to see how forensic science can become a true science in that environment. Science flourishes when free and independent; only then can the tools and technology that it creates be truly reliable.<sup>383</sup>

The American Association for the Advancement of Science concurred, also stressing that independence “cannot be overstated” and expressing concern about the “inherent conflict of interest in having law enforcement overseeing the work of forensic labs on which police and prosecutors rely to win and defend convictions.”<sup>384</sup> The American Academy of Forensic Science also opposed the formation of an Office of Forensic Science within DOJ.<sup>385</sup> Instead of heeding this advice, the Attorney General subsequently appointed a prosecutor instead of a scientist a forensic science to head the working group within the DOJ.<sup>386</sup>

These recent events should be put in context. The NAS 2009 report recommended the creation of an independent federal entity (the National Institute of Forensic Sciences) to oversee the field, including a research agenda.<sup>387</sup> If adopted, this proposal would have wrest control of forensic science from law enforcement. The report provided the following justification: Some federal entities were “too wedded” to the status quo and “have failed to pursue a rigorous research agenda to confirm the evidentiary reliability of methodologies used in a number of forensic science disciplines.”<sup>388</sup> As a result, these “agencies are not good candidates to oversee the overhaul of the forensic science community.”<sup>389</sup>

---

<sup>383</sup> Sunita Sah et al., Observations, *We Must Strengthen the “Science” in Forensic Science*, SCI. AM., May 8, 2017.

<sup>384</sup> Spencer S. Hsu, *Science Organizations Renew Call for Independent U.S. Committee on Forensics*, WASH. POST, June 29, 2017 (“The association linked the problem to what it described as an inherent conflict of interest in having law enforcement overseeing the work of forensic labs on which police and prosecutors rely to win and defend convictions.”)

<sup>385</sup> Message From the AAFS President, American Academy of Forensic Science, April 2017.

<sup>386</sup> See Pam Levy, *Sessions’ New Forensic Science Adviser Has a History of Opposing Pro-Science Reforms*, Mother Jones, Aug. 10, 2017 (“But Attorney General Jeff Sessions has resisted efforts to rein in forensic science and hold it to higher standards. And this week, he appointed a senior adviser on forensics who has a history of opposing reforms that would bring more accountability and scientific rigor to forensic crime labs and expert testimony.”).

<sup>387</sup> NAS FORENSIC REPORT, *supra* note 21, at 19 (Recommendation 1(c): “promoting scholarly, competitive peer-reviewed research and technical development in the forensic science disciplines”).

<sup>388</sup> *Id.* at 18.

<sup>389</sup> *Id.* There is little question that the committee was referring to National Institute of Justice and the FBI Laboratory. The report noted that, although both had provided “modest leadership” in forensic science, “neither entity has recognized, let alone articulated, a need for change or a vision for achieving it.” *Id.* at 16. The report also stated: “Neither has the full



There is little question that the NAS was referring to National Institute of Justice and the FBI Laboratory. The report noted that, although both had provided “modest leadership” in forensic science, “neither entity has recognized, let alone articulated, a need for change or a vision for achieving it.”<sup>390</sup> Consequently, “advancing *science* in the forensic science enterprise is not likely to be achieved within the confines of DOJ.”<sup>391</sup> In fact, law enforcement had manipulated science by shaping the research agenda, limiting access to data, attacking experts who disagreed with its positions, and “spinning” negative reports.<sup>392</sup>

When Congress did not authorize the creation of the National Institute of Forensic Sciences, DOJ, to its credit, established the NCFS. Most importantly, independent scientists were appointed to the commission.<sup>393</sup> Now placing science back under DOJ is a major and unjustified retreat from science.

## V. CONCLUSION

In this article, I explained how the judiciary’s failure to fulfill its gatekeeper role can be traced back to its refusal to demand and properly evaluate foundational research, i.e., *Daubert’s* first factor (empirical testing). This failure has been systemic. Flawed forensic techniques such as bite mark analysis, microscopic hair comparisons, arson evidence, and comparative bullet lead analysis were routinely admitted into evidence without foundational research. In addition, firearms/toolmark and fingerprint examiners repeatedly presented overstated and misleading conclusions. I also argued that the justice system may be institutionally incapable of applying *Daubert* in criminal cases because it does not have access to independent scientific expertise on an ongoing basis, and I endorsed the NCFS and PCAST recommendation that NIST should be tasked with this responsibility.

Even if an independent scientific review is not institutionalized, PCAST, NCFS, and AAAS have provided guidance for courts dealing with admissibility challenges. First, the flawed techniques discussed in this article should be

---

confidence of the larger forensic science community. And because both are part of a prosecutorial department of the government, they could be subject to subtle contextual biases that should not be allowed to undercut the power of forensic science.” *Id.* Consequently, “advancing *science* in the forensic science enterprise is not likely to be achieved within the confines of DOJ.” *Id.* at 18.

<sup>390</sup> *Id.* at 16. The Report also stated: “Neither has the full confidence of the larger forensic science community. And because both are part of a prosecutorial department of the government, they could be subject to subtle contextual biases that should not be allowed to undercut the power of forensic science.” *Id.*

<sup>391</sup> *Id.* at 18.

<sup>392</sup> See Giannelli, *supra* note 344.

<sup>393</sup> Having served on the NCFS, I believe that there should have been more.

excluded. If used at all, bite mark analysis should be limited to exclusions and perhaps to closed universe situations.<sup>394</sup> For hair analysis, mitochondrial DNA analysis is far superior to microscopy. Arson evidence should comport with NFPA 921 and the AAAS report. As noted above, the FBI has abandoned comparative bullet lead analysis.

Second, courts should focus, as *Daubert* requires, on foundational research: “[E]xperience, judgment, good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) cannot substitute for actual evidence of foundational validity and reliability.”<sup>395</sup> The NCFS concurred.<sup>396</sup>

Third, subjective methods can be empirically tested. Such research has been conducted. PCAST identified studies in fingerprint and firearms identification that meet stringent standards.<sup>397</sup> These studies show an error rate, which should be presented to the jury.<sup>398</sup> However, more than one study is needed.

Fourth, in ruling on admissibility in firearms/toolmark and fingerprint examination cases, courts should appreciate that there has been a “lost decade” (or two) during which rigorous research was not conducted.<sup>399</sup> Instead, the disciplines

---

<sup>394</sup> See, e.g., *State v. Lambright*, 2014 Tenn. Crim. App. LEXIS 5 (Tenn. Crim. App. Jan. 7, 2014) (“Dr. Tabor said that, considering the number of teeth that the victim’s sister had, she would not have been capable of producing the bite mark found on the victim’s nose and upper lip. It was Dr. Tabor’s expert medical opinion that a two-year-old was not capable of producing the nature, severity, number, and orientation of bites sustained by the victim.”).

<sup>395</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 5-6.

<sup>396</sup> See *supra* text accompanying note 350 (emphasizing the importance of published peer review research).

<sup>397</sup> See *supra* text accompanying note 336 (noting that closed set studies are not sufficiently robust).

<sup>398</sup> If examiners claim that there is no error rate, they should be required to explain why not. See WHITE HOUSE PCAST REPORT, *supra* note 33, at 19 (“In testimony, examiners should always state clearly that errors can and do occur, due both to similarities between features and to human mistakes in the laboratory.”).

<sup>399</sup> This depends on when the clock started ticking. *Daubert* was decided in 1993. In 1995, the first challenged to handwriting testimony was decided. *United States v. Starzeczyzel* 880 F. Supp. 1027 (S.D.N.Y. 1995). As noted above, the *Williamson* case on microscopic hair analysis was decided the same year. See *supra* text accompanying note 79-80. *Llera Plaza* (fingerprints) was decided in 2002. See *supra* text accompanying note 283-85. *Green* (firearms identification) in 2005. See *supra* text accompanying note 283-85. 237-30. Moreover, during this period numerous courts restricted the use of handwriting identification. See, e.g., *United States v. Hines*, 55 F. Supp. 2d 62, 67 (D. Mass. 1999) (holding that expert testimony concerning the general similarities and differences between a defendant’s handwriting exemplar and a stick up note was admissible but not the specific conclusion that the defendant was the author).



examined in this article vigorously resisted the views of independent scientists. For example, the wrongful execution of Cameron Todd Willingham which triggered numerous scientific reviews was not enough to persuade the Texas Fire Marshal Office that its evidence was flawed,<sup>400</sup> and it took a serendipitous event — the Madrid train bombing — to provoke fingerprint research.

In short, a “Catch-22” situation: only the federal government has the resources to fund the needed independent research, but there was no incentive to do so as long as evidence continued to be admitted without proper limitations. Until there are more scientifically sound studies that have been published and peer-reviewed by independent scientists, courts should follow the approach adopted in *United States v. Glynn*,<sup>401</sup> which permitted the expert to testify only that it was “more likely than not” that recovered bullets and cartridge cases came from a particular weapon.<sup>402</sup>

Fifth, the presentation of expert testimony needs to be controlled. Once again, PCAST made several recommendations, including:

Statements suggesting or implying greater certainty are not scientifically valid and should not be permitted. In particular, courts should never permit scientifically indefensible claims such as: “zero,” “vanishingly small,” “essentially zero,” “negligible,” “minimal,” or “microscopic” error rates; “100 percent certainty” or proof “to a reasonable degree of scientific certainty;” identification “to the exclusion of all other sources;” or a chance of error so remote as to be a “practical impossibility.”<sup>403</sup>

The NCFS also recommended against the use of the phrase “reasonable degree of scientific certainty”<sup>404</sup> and the 2009 NAS report criticized the use of “zero error rates” and claims of infallibility.<sup>405</sup> The recent AAAS fingerprint report found no scientific justification for statements of “identity” or “practical certainty” and cautioned against the use of terms such as “match,” “identification,” and “individualization.”<sup>406</sup>

Unfortunately, there is little reason to believe that examiners will give up their claims that there is a scientific foundation for their discipline. A subjective

---

<sup>400</sup> See *supra* text accompanying note 140.

<sup>401</sup> 578 F. Supp. 2d 567 (S.D.N.Y. 2008).

<sup>402</sup> *Id.* at 575.

<sup>403</sup> WHITE HOUSE PCAST REPORT, *supra* note 33, at 19.

<sup>404</sup> See *supra* text accompanying note 254.

<sup>405</sup> See *supra* text accompanying note \_\_.

<sup>406</sup> AAAS FINGERPRINT REPORT, *supra* note 34.



method without a meaningful protocol can hardly claim to be a science. This is not a new issue, a editorial in the prestigious scientific journal, *Science*, entitled “Forensic Science: Oxymoron?” and written by the editor-in-chief, made the same point fifteen years ago.<sup>407</sup> Similarly, the 2009 NAS report commented: “The law’s greatest dilemma in its heavy reliance on forensic evidence . . . concerns the question of whether — and to what extent — there is *science* in any given forensic science discipline.”<sup>408</sup> After *Daubert* hearings, one court “very quickly concluded that whatever else ballistics identification analysis could be called, it could not fairly be called ‘science.’”<sup>409</sup> The same is true of fingerprint examinations.

Courts should also guard against attempts to introduce “science” through the backdoor by means of circumlocutions such as firearms or fingerprint identifications are subjective techniques “based on science.” This is misleading. Many things are “based on science” — e.g., riding a bike, throwing a curve ball, and flying a kite.

Sixth, proficiency testing issues will continue to be litigated. These tests have long been suspect. They are not conducted blind and are not challenging.<sup>410</sup>

---

<sup>407</sup> Donald Kennedy, Editorial, *Forensic Science: Oxymoron?*, 302 *SCIENCE* 1625 (2003). The editorial discussed the cancellation of a National Academies project designed to examine various forensic science techniques, including fingerprinting, because the Departments of Justice and Defense insisted on a right of review that the Academy had refused to other grant sponsors.

<sup>408</sup> NAS FORENSICS SCIENCES REPORT, *supra* note 21, at 9.

<sup>409</sup> *United States v. Glynn*, 578 F. Supp. 2d 567, 570 (S.D.N.Y. 2008). *See also Starzecpyzel*, 880 F. Supp. at 1038, 1041 (“[F]orensic document examination, despite the existence of a certification program, professional journals and other trappings of science, cannot, after *Daubert*, be regarded as ‘scientific . . . knowledge.’”; “[W]hile scientific principles may relate to aspects of handwriting analysis, they have little or nothing to do with the day-to-day tasks performed by [Forensic Document Examiners] . . . . [T]his attenuated relationship does not transform the FDE into a scientist.”).

<sup>410</sup> For example, a fingerprint examiner from New Scotland Yard testified in one case that the FBI proficiency tests were deficient: “It’s not testing their ability. It doesn’t test their expertise. I mean I’ve set these tests to trainees and advanced technicians. And if I gave my experts these tests, they’d fall about laughing.” *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 565 (E.D. Pa. 2002). The district court agreed, noting that “the FBI examiners got very high proficiency grades, but the tests they took did not. . . . [O]n the present record I conclude that the proficiency tests are less demanding than they should be.” *Id.* at 558. Similarly, in a trial involving handwriting comparisons, the court wrote:

There were aspects of Mr. Cawley’s testimony that undermined his credibility. Mr. Cawley testified that he achieved a 100% passage rate on the proficiency tests that he took and that all of his peers *always* passed their proficiency tests. Mr. Cawley said that his peers *always* agreed with each others’ results and *always* got it right. Peer review in such a “Lake Woebegone” environment is not meaningful.

*United States v. Lewis*, 220 F. Supp. 2d 548, 554 (S.D.W. Va. 2002). *See supra* note 307

The President of Collaborative Testing Services told the NCFS “during its seventh meeting (August 10, 2015) that he has been under commercial pressure to make proficiency tests easier.”<sup>411</sup>

---

(discussing fingerprint proficiency testing).

<sup>411</sup> National Commission on Forensic Science, Views of the Commission, Optimizing Human Performance in Crime Laboratories through Testing and Feedback, May 27, 2016, at <https://www.justice.gov/ncfs/file/864776/download>.

85 Fordham L. Rev. 619

Fordham Law Review  
November, 2016

Symposium  
Criminal Behavior and the Brain: When Law and Neuroscience Collide  
Erin Murphy<sup>a1</sup>

Copyright © 2016 by Fordham Law Review; Erin Murphy

## NEUROSCIENCE AND THE CIVIL/CRIMINAL DAUBERT DIVIDE

### INTRODUCTION

Advances in neuroscience have dramatically expanded our knowledge of the brain and how it operates. Although many mysteries remain, the early architectures of our understanding have already left impressions on the legal system. Neuroscientific evidence has been offered to support claims by litigants in both civil and criminal cases, ranging from broad-based generalities (such as “juvenile brains are generally immature in these ways”) to individualized opinions (such as “this defendant lacked the cognitive capacity to control this behavior”).

As such evidence trickles into the courts, scholars have debated the scientific foundation of such claims, the scope of their applicability, and whether such evidence has met some threshold of reliability imposed before courts and fact-finders ought to accept them.<sup>1</sup> But most scholarly treatments of neuroscientific proof overlook a more fundamental question regarding evidentiary admissibility: What impact will the standard applied to determine admission--both de jure and de facto--have on the rate of acceptance of this new evidence? History suggests that, when it comes to proffers of scientific evidence, civil and criminal proceedings are not in fact created equal. Moreover, the application of evidentiary standards varies widely, and constitutional oversight of evidentiary rules is, for litigants other than a criminal defendant, somewhere between threadbare and nonexistent.

\*620 This Article thus speculates on the course of neuroscience-as-proof with an eye toward the actual admissibility standards that will govern the acceptance of such evidence by courts, not just as a matter of formal law but also as a function of historical custom. Given the legal system’s spotty record with scientific evidence--which is to say, both the demonstrated willingness of the system to admit unproven “science” or to exclude evidence despite a seemingly adequate scientific foundation--the trajectory of neuroscience in the courts cannot be predicted simply by asking about its scientific legitimacy in the abstract. Rather, an observer must ponder whether patterns of admissibility long evident in criminal and civil courts will persevere with respect to neuroscientific proof.

One clarification is warranted. Throughout this Article, I use the phrases “novel neuroscience” and “novel neuroscientific evidence.” Capturing precisely what is meant by “neuroscience,” much less “novel neuroscience,” can often prove more elusive than seems at first glance.



I generally follow Professor Nita Farahany's approach, which prefers the word "neurobiological" to capture "evidence about the study of the brain and the nervous system," which includes "claims about the 'normal' brain, abnormal brain, effects on neurotransmitters, brain structure, function, and genetic contributions to neurological functioning and structure."<sup>2</sup> Professor Farahany's definition also broadly encompasses evidence based on imaging techniques (such as CT or MRI), as well as findings drawn from interviews (intended to elicit, for instance, whether a person had a brain injury) or psychological assessments.<sup>3</sup>

I further circumscribe this category to "novel neuroscience." By this, I mean to exclude relatively noncontroversial uses of neuroscience, such as those that show an undisputed physical insult or injury to the brain, or its fairly noncontroversial consequence, like a car accident that results in visible damage to a portion of the brain affecting speech, where the injured person developed precisely that expected speech impairment. I also intend to exclude assessments that have only remote connection to the physical condition of the brain, such as psychological assessments that have no connection to any observed physiological conditions. In short, I mainly intend to speak to precisely what the phrase suggests: novel or cutting-edge methods--whether scan-based or assessment-based--that purport to link a finding about the structure or physiological function of the brain to a manifested behavior, cognitive power, or psychology. Moreover, this Article considers the likely treatment of novel neuroscientific evidence when offered in courts at this moment in scientific understanding; in other words, it does not assume any game-changing breakthroughs on what may reliably be proven.

Part I begins by recounting the historical divide between civil and criminal courts with respect to the treatment of novel scientific evidence. Part II then explores, both by examining current trends and predicting future trajectories, \*621 whether this pattern of differential treatment is likely to endure as courts begin to confront the admissibility of novel neuroscience.

## I. DAUBERT'S TWO FACES: CIVIL V. CRIMINAL

The formal standard for admission of expert evidence may, as a matter of formal law, be the same in civil and criminal cases. But in practice, both scholars and litigants have observed that the application of that standard varies markedly. The conventional wisdom holds, and empirical studies support, that evidence proffered by plaintiffs in civil cases receives harsh scrutiny for reliability, whereas evidence proffered by prosecutors in criminal cases typically gets a free pass. But, as explained in this part, this disparity is rarely observed directly because apart from a couple of exceptions--most notably fire science and handwriting analysis--the type of evidence offered by civil litigants has little overlap with that offered by criminal prosecutors.

### A. Background

When announced by the U.S. Supreme Court in 1993, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*<sup>4</sup> was heralded as a watershed moment in the treatment of scientific evidence.<sup>5</sup> In its opinion, the Court displaced the longstanding *Frye v. United States*<sup>6</sup> "general acceptance" test ("the *Frye* test") as the standard for evidentiary admissibility. With the Court's opinions in *General Electric Co. v. Joiner*<sup>7</sup> and *Kumho Tire Co. v. Carmichael*<sup>8</sup> that quickly followed, the Supreme Court seemed to erect an entirely new and more rigorous test for admissibility intended to stem the perceived epidemic of "junk science" that had overtaken the courts.<sup>9</sup>

But even in the midst of this celebration, suspicions began circulating that *Daubert*'s professed commitment to rigorous examination of evidence offered in civil cases--like the one in which the ruling was announced--would not extend to its criminal brethren. For instance, the opinion itself, which talked breathlessly about the scientific ideal of "reliability" in ways later criticized by philosophers of science,<sup>10</sup> conspicuously omitted any reference to the forensic sciences that routinely arose in criminal courts. \*622 Then, on remand, Judge Alex Kozinski of the Ninth Circuit, palpably bristling at the "daunting" task of acting as an arbiter of scientific reliability,<sup>11</sup> took pains to exempt "[f]ingerprint analysis, voice recognition, DNA fingerprinting and a variety of other scientific endeavors closely tied to law enforcement" from *Daubert*'s strictures, setting up a de facto divide between civil and criminal *Daubert*.<sup>12</sup>

In the years since the *Daubert* trilogy--which also witnessed amendments to Rule 702 of the Federal Rules of Evidence that either codified or enhanced its standards, depending on whom you ask<sup>13</sup>--the debate over *Daubert*'s impact has continued. Such findings have political and not just legal significance because in both civil and criminal cases, the methods and techniques most vulnerable to *Daubert* scrutiny, as judged by scientific standards, tend to be offered by only one side in the litigation. And in fact, those parties even sit on the same side of the courtroom: prosecutors in criminal cases and plaintiffs' attorneys in civil cases. That is, plaintiffs' attorneys, such as in toxic tort or personal injury cases, often rest their proof on medical or scientific findings that are readily challenged as unreliable by defendants.<sup>14</sup> Similarly, prosecutors in criminal cases routinely offer evidence based on methods like fingerprinting, hair and fiber analysis, or pattern matching (like ballistics or bite marks), notwithstanding reliable indicators that such evidence is in fact wholly lacking in scientific support.<sup>15</sup>

Even though "*Daubert* ostensibly applies in the same way in criminal and civil cases, social scientists have increasingly raised the issue whether courts, in fact, apply *Daubert* more lackadaisically in criminal trials-- especially in regard to prosecution evidence."<sup>16</sup> Given that the proponents of vulnerable scientific evidence tend to hew to one side, the degree to which *Daubert* works to exclude such science carries important repercussions for measurements of plaintiff and prosecutorial success. Thus, multiple empirical studies have endeavored to answer precisely whether *Daubert* has, in fact, served its role of precluding junk science while admitting reliable, even if cutting-edge or novel, techniques.<sup>17</sup>

Generally speaking, these studies themselves divide between civil and criminal cases. And they seem to reaffirm, albeit imperfectly, the intuition of litigants and those familiar with the justice system: "civil defendants win \*623 their *Daubert* reliability challenges to plaintiffs' proffers most of the time, and ... criminal defendants virtually always lose their reliability challenges to government proffers."<sup>18</sup> In short, "civil defendants have benefited greatly from *Daubert* but ... criminal defendants have not."<sup>19</sup>

One iconic comparison was conducted by Professor Michael Risinger in 2000. He looked at over 1,600 citations to *Daubert* by American state and federal courts, in a period from 1993 to 1999, and compared that to a reference set of opinions citing *Frye* in the six years prior to *Daubert*.<sup>20</sup> He found that post-*Daubert*, courts excluded plaintiffs' proffered evidence at high rates, even while granting plaintiffs' requests to exclude defense evidence at much lower rates.<sup>21</sup> On the criminal side, he found that defense challenges to prosecution evidence infrequently succeeded, even while prosecution challenges to defense evidence had roughly the same success rate as that of civil defendants.<sup>22</sup>

Professor Risinger's findings have been replicated by others using an array of approaches.<sup>23</sup> Those findings show that in the civil context, generally speaking, "studies show that after *Daubert*, parties challenged the admissibility of evidence more frequently, and judges scrutinized evidence more carefully, excluding a greater proportion of it."<sup>24</sup> In contrast, in the criminal context, one major review found that questioned experts tended to testify for the prosecution, and "the *Daubert* decision did not impact on the admission rates of expert testimony at either the trial or the appellate court levels."<sup>25</sup>

Some observers might wonder whether these findings simply reflect the relative substantive merit of evidence offered by civil plaintiffs versus criminal prosecutors. Indeed, if it simply is the case that prosecutors offer robust, reliable techniques, whereas civil plaintiffs tend to offer novel, untested methods, then these findings simply show that the standard is performing as expected. But regardless of the merits of plaintiffs' \*624 evidence--which is a subject of some debate--that conjecture does not bear out with respect to prosecutorial evidence. Consider that nearly all of the common forensic techniques offered by prosecutors, and routinely admitted by courts, have been repeatedly denounced as lacking in any scientific basis.<sup>26</sup> Most prominently, a 2009 National Academy of Sciences report observed that

[w]ith the exception of nuclear DNA analysis, ... no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source .... The simple reality is that the interpretation of forensic evidence is not always based on scientific studies to determine its validity.<sup>27</sup>

Indeed, some criminal courts admitting forensic evidence despite defense challenges to reliability have expressly conceded that the proposed conclusions lack any scientific basis in data, methods, or statistical significance--and yet nonetheless embraced them citing nothing more than their longstanding pedigree.<sup>28</sup>

In sum, commentators, scholars, and courts themselves seem to acknowledge that there exists a *Daubert* double standard. Professor Jane Moriarty has further intimated that this double standard is not just the product of incompetence or lack of understanding.<sup>29</sup> She notes that

[i]n civil cases, courts seem quite up to the task of evaluating microbiology, teratology, and toxicology evidence, discussing both science and statistics with plenty of acumen. Yet when it comes to evaluating the shortcomings of lip prints and handwriting, courts are unable to muster the most minimal grasp of why a standardless form of comparison might lack evidentiary reliability or trustworthiness.<sup>30</sup>

This intuition is perhaps bolstered by efforts to expressly enshrine the distinction. In the wake of *Daubert*, federal lawmakers circulated a bill to exempt criminal evidence from the proposed codification of the *Daubert* test, but their efforts failed.<sup>31</sup> That suggests that political actors, or at least some legislators, would expressly aim to lower the bar of reliability for evidence admitted in criminal cases. But whether de facto or de jure, the bottom line seems that, whatever *Daubert*'s bark, it tends to bite only in civil cases.

#### \*625 B. Exceptions

The disparate treatment of proffered scientific evidence in the civil and criminal context is easily masked in part because the disciplines relied upon in each context diverge so sharply. In the civil context, experts tend to offer opinions about causal factors of injury or illness.<sup>32</sup> In the criminal context, by contrast, experts tend to be less concerned with causation and more focused on identification.<sup>33</sup> The civil cases are littered with examples of doctors, epidemiologists, and social scientists offering medical and mechanical explanations,<sup>34</sup> whereas the criminal cases consist largely of devoted forensic analysts--often police department employees--discussing methods like fingerprinting, trace evidence identification, handwriting analysis, and the like.<sup>35</sup>

Even scientific disciplines that may, on the surface, appear to apply in both civil and criminal contexts do not upon closer examination. For instance, DNA typing is a scientific technique that obviously carries great import for criminal cases as an identification method, and it is also easy to imagine that it might be relevant in a civil case involving genetic testing of some kind. But, for reasons that are too complex to detail in this Article, the methods, instrumentation, and interpretive difficulties of DNA testing in each context are in fact quite different.<sup>36</sup> Even DNA testing in civil parentage cases--the closest analogue to the criminal context--diverges significantly from the kinds of reliability challenges that arise in criminal forensic testing. To give just one example, parentage testing always involves controlled quality and quantity samples taken from known individuals (the putative parents or the child), whereas forensic testing focuses on crime scene samples from unknown persons collected in uncontrolled conditions that may be of low quality or quantity.<sup>37</sup>

There are, however, two disciplines that form an area of overlap between civil and criminal cases and thus might directly surface the conflict between civil and criminal admissibility standards. Specifically, fire investigation is relevant for both criminal arson and civil insurance cases, and handwriting analysis is pertinent for both criminal cases and civil cases. These two areas thus provide good source material against which to test the thesis that courts apply admissibility standards more strictly in civil cases (to evidence offered by plaintiffs) than in criminal cases (to evidence offered by prosecutors).

\*626 A 2013 article by Professor Julie Seaman probed a version of this question.<sup>38</sup> Professor Seaman sought to answer whether the same discipline (fire science or handwriting analysis) received different treatment depending on the kind of case (civil versus criminal).<sup>39</sup> In a review that she conceded faced some methodological challenges,<sup>40</sup> she made some interesting findings. In short:

Comparing the admission and exclusion percentages in criminal and civil cases, then, it is apparent that the disparity seen in the handwriting cases is not evident in the fire cause and origin cases. In the handwriting



cases, prosecution evidence was admitted in nearly 90% of the criminal cases, whereas on the civil side it was admitted (or at least not excluded) in fewer than 40% of cases. In contrast, the admission rates for expert testimony in the fire cases hovered close to 75% for both criminal and civil cases.<sup>41</sup>

On its face, these findings present a conflicting image.<sup>42</sup> But examined more closely, they reaffirm and deepen the initial underlying premise: it depends as much on the offering party as it does on the type of case. In criminal cases involving fire science, the prosecution (the favored party) tends to offer the evidence, and so we would expect high rates of admission. In civil cases, however, it is not only plaintiffs that offer this evidence but rather civil defendants as well; fire science experts tend to be used by defendant-insurers who seek to defend against claims lodged by plaintiff-insureds.<sup>43</sup> Thus, if the evidence is admitted in civil cases at high rates, it may very well be because it is offered by the favored party in those cases--the defendant.

By contrast, the cases involving handwriting analysis fit the more typical picture. Handwriting analysis tends to be offered by the prosecution in the criminal context and by plaintiffs in the civil context.<sup>44</sup> And again, Seaman found that in criminal cases, the admission rate was around 90 percent, whereas the exclusion rate in civil cases was roughly 64 percent.<sup>45</sup> Importantly, in looking at the qualitative language used in these cases to discuss the admission or exclusion determination, Seaman found marked variation in the perspective of judges:

Whereas in criminal cases, for the most part, the global field of questioned document analysis is one with a long history, tested in the crucible of the \*627 adversarial process and relied upon by law enforcement and overwhelmingly approved by courts, in civil cases the field is peopled by unqualified charlatans who use untested methodologies to offer unreliable opinions that are not helpful to juries, which are perfectly capable of comparing handwriting samples on their own.<sup>46</sup>

In short, although handwriting analysis or fire science evidence arises in both the criminal and civil contexts, when it comes to judging the admissibility of the proffered evidence, each discipline's rate of success follows the same pattern of admission and exclusion apparent from studies about the rigor of *Daubert* when it comes to nonoverlapping fields. When faced with evidence offered by prosecutors or civil defendants, courts tend to take a generous approach, whereas even the same kind of evidence offered by civil plaintiffs is met with great skepticism.

## II. THE NEW KID ON THE BLOCK: NEUROSCIENTIFIC EVIDENCE

Given the conventional wisdom, borne out by empirical study, that *Daubert* bites in civil cases but merely barks in criminal ones, how might we expect courts to treat the impending onslaught of neuroscientific evidence? Like handwriting analysis and fire investigation, novel neuroscience creates a point of tension because it can arise in both categories of cases and be introduced by either side in a dispute. Specifically, novel neuroscientific methods, such as those used to detect closed brain injuries or subtle cognitive, emotional, or psychological conditions, have cross-applications that make them more like handwriting analysis than like side-specific methods such as idiopathic mesothelioma or bite marks. If novel neuroscience extends beyond its present reach--most commonly to capital criminal defendants and to a lesser extent to civil plaintiffs--and becomes part of the prosecutorial and perhaps even civil defendants' arsenal,<sup>47</sup> what will happen? Novel neuroscientific evidence may present the law with the direct point of conflict that it has henceforth averted: the context and side-specific treatment of scientific evidence, whether civil versus criminal cases or plaintiffs and prosecutors versus defendants. And from that conflict, observers may gain a clearer sense of the successes and failures of our evidentiary admissibility standard.

What will be the result of this point of conflict? Will admissibility standards operate to preclude novel neuroscientific evidence, and, if so, in what kinds of cases and by which parties? Will neuroscience admissibility patterns reflect the same political story recounted above, or will they cleave between prosecutorial evidence and plaintiffs' evidence as hinted at by the findings in the handwriting example? Will pressure to reconcile these admissibility decisions result in the forging of some new equilibrium? It is \*628 too early to know, but the remainder of this Article will consider current trends and explore

several possibilities.

### A. Current Trends

In both civil and criminal cases, neuroscientific evidence commonly has been introduced to support noncontroversial findings such as structural damage or major brain injury, easily readable on a standard CT or MRI scan.<sup>48</sup> Although such findings are not without challenge, they tend to be relatively noncontroversial. The novel neuroscientific proof of interest to this Article, however, is that which relies on contested questions such as the degree to which conclusions can be drawn about a single individual from aggregated group data,<sup>49</sup> the relationship between cause and effect, the absence of baseline data about a subject's brain prior to trauma,<sup>50</sup> or the ascertainment of disputable injuries or abnormalities.<sup>51</sup> These developments raise some degree of alarm on the part of scientists when applied in a context of categorical decision making--such as the recent Supreme Court decision citing neuroscience about juveniles as a basis for a wholesale prohibition on the death penalty or mandatory life without parole for that group<sup>52</sup>--but the most contested use continues to be to support findings individualized to a specific person.

In civil cases, plaintiffs most commonly offer novel neuroscientific evidence for one of three different purposes: (1) to show brain injury, in particular closed head injuries; (2) to prove the existence of toxic encephalopathy or other chemical sensitivities; and (3) for lie detection.<sup>53</sup> In criminal cases, novel neuroscientific evidence is typically admitted at the request of the defendant in support of arguments to mitigate punishment, most often in serious sentencing hearings like capital cases.<sup>54</sup>

\*629 Thus far, courts' response to neuroscientific evidence when offered for these purposes has been tentative and inconsistent. Courts have shown the greatest enthusiasm for admitting evidence offered by capital defendants seeking to fight a sentence of death by showing brain conditions that mitigate their criminal responsibility. In this context, courts have admitted neuroscientific evidence to bolster claims of behavioral or emotional disorders,<sup>55</sup> the absence of a culpable mental state or evidence of insanity,<sup>56</sup> and diminished cognitive capacity.<sup>57</sup> But it is only the use of neuroscientific evidence in the mitigation phase that has become genuinely common--so common, in fact, that appellate judges have even found that failure to investigate neuroscientific explanations for behavior constitutes ineffective assistance of counsel.<sup>58</sup>

In civil cases, judges have shown greater recalcitrance about admitting novel neuroscientific proof, although there are occasional exceptions. For instance, although courts routinely admit established technologies like CT, PET, and MRI scans as proof of major structural damage to a brain, they have not always welcomed such evidence when offered to prove the existence or cause of minor closed-head brain trauma (often abbreviated "TBI" for "traumatic brain injury").<sup>59</sup> There are a handful of examples to the contrary, but courts still typically exhibit significant reservation about allowing in such evidence. When it comes to cutting-edge methods like QEEG or SPECT,<sup>60</sup> as well as novel findings such as toxic encephalopathy<sup>61</sup> or lack of truthfulness (lie detection),<sup>62</sup> courts have overwhelmingly rejected such proffered evidence as unreliable.

Of course, broader applications of neuroscientific evidence are easily imaginable. As succinctly laid out by one group of authors, neuroscientific evidence could answer questions as wide ranging as:

[I]s this person responsible for his behaviour? What was this person's mental state at the time of the act? How much capacity did this person have to act differently? What are the effects of addiction, adolescence or advanced age on one's capacity to control behaviour? How competent is this person? What does this person remember? How accurate is this person's memory? What are the effects of emotion on memory, behaviour and motivation? Is this person telling the truth? In how much pain is this person? How badly injured is this person's brain?<sup>63</sup>

Although there are occasional examples of courts admitting novel neuroscientific evidence in support of some of these outlier propositions, in \*630 general courts find such evidence unreliable under a *Daubert*, *Frye*, or other pertinent standard.<sup>64</sup>

In sum, courts in civil cases tend to reject novel neuroscientific evidence unless it supports fairly solid-seeming claims of

traumatic brain injury, and in criminal cases, courts express similar reluctance to admit evidence unless it is offered as mitigation evidence. But when so offered, and in particular as capital mitigation, courts tend to take a more permissive view of admissibility.

### **B. Future Directions**

What do these early patterns of neuroscientific admissibility patterns predict for the future? This part, of necessity, constitutes pure speculation. But for the sake of argument, let us presume two things. First, assume that in the near term, claimants will continue to proffer neuroscientific evidence, and courts will continue to face challenges on the basis of scientific validity. In other words, do not expect that these early defeats will dissuade litigants from continued efforts to utilize neuroscientific evidence. Second, assume that the state of the science continues to improve. Methods become more robust and technologies advance. Thus, while still fraught, such findings refine incrementally in terms of specificity and sensitivity. What might we expect the arc of admissibility to look like, knowing what we do about the courts' track record when it comes to novel or unproven scientific techniques?

#### **1. A Ban: Novel Neuroscience Goes the Way of the Polygraph**

One possibility is that neuroscientific evidence will continue to meet broad resistance by courts, which will remain skeptical of its reliability and mindful of the numerous cautions sounded by scientists who aim to curb efforts of overclaiming.<sup>65</sup> Under this view, the current trends of excluding novel neuroscientific evidence in the vast majority of civil and criminal cases will continue, with perhaps a small pocket of admission when offered by defense in mitigation proceedings (more on that later). The enthusiasm of proponents of neuroscience will thus ultimately be checked by courts, which will strictly apply the standards of evidentiary admissibility and deem most methods insufficiently reliable.

Evidence of this kind of skepticism is already apparent in existing civil cases, where plaintiffs, generally speaking, have failed when proffering in evidence a wide array of uses of novel neuroscience.<sup>66</sup> It is also to some degree evident in the criminal cases, where defendants outside of the mitigation context tend to meet similar skepticism. Indeed, fears about prosecutorial overreaching, the usurpation of the jury function, and "trial by \*631 machine" might further work to stem the tide in criminal cases. Thus, going against the conventional practice of imposing stricter admissibility tests on plaintiffs than on prosecutors, courts might simply reject novel neuroscience altogether.

Such a result would not be unprecedented. For instance, when ordinary lie detector tests first came to market, there existed a similar fervor that such tests offered a scientifically certain means of resolving law's recurring problem of assessing human credibility.<sup>67</sup> But the tool proved quite useful to defendants because it offered "scientific" validation of their honesty.<sup>68</sup> In fact, it was offered for just that purpose in *Frye*, the landmark case that announced the reliability standard that dominated American law for decades, and the court ruled it insufficiently reliable on the grounds that it had not yet gained general acceptance.<sup>69</sup>

But lie detection methods did not fade. Defense lawyers continued to argue their applicability for purposes other than formal admission as evidence.<sup>70</sup> Nevertheless, in the wake of *Frye*, "[t]he vast majority of courts maintained a per se inadmissibility rule."<sup>71</sup> Then, as polygraph technology improved, and the Supreme Court laid down the *Daubert* standard, there was a resurgence in hope that the polygraph might return to court.<sup>72</sup> By this point, law enforcement had routinely relied on polygraphs for making charging and other decisions, so it seemed that the method might gain greater favor. Indeed, polygraph machines arguably have a stronger scientific foundation than numerous other forensic methods--such as bite mark or tool mark matching--that have earned widespread acceptance in criminal courts.<sup>73</sup> Yet even when revisited in the wake of *Daubert*, polygraphs still could not make it into court.<sup>74</sup> To be fair, some of those judgments turned on concerns other \*632 than reliability, such as undue prejudice to the jury.<sup>75</sup> But courts seemed to have gained familiarity with the idea that polygraphs had no place in evidence and did not miss them. Whatever the reason, continued exclusion was the path of least resistance.



The same kind of story easily could play out with regard to novel neuroscience. Like those initial polygraphs, the successful proponents of this evidence tend, at this time, to be criminal defendants.<sup>76</sup> And like polygraphs, novel neuroscience raises concerns about displacing the function of juries and confusing the fact-finder; indeed, proposed uses of novel neuroscience include lie detection, superseding the polygraph.<sup>77</sup> If courts deem such evidence insufficiently reliable, perhaps even bolstered by the findings of courts in civil cases where such evidence is offered by plaintiffs, then exclusion may become the default in much the same way it has with lie detection testing, notwithstanding improvements to the technology or error rate over time. In such a scenario, novel neuroscience may go the way of the polygraph machine--perhaps operating in the background to inform the choices of actors in the system, but never taking its place as full-bodied evidence in court, regardless of any gains in the reliability of specific uses.

## 2. The Same Old Story: Prosecutors' Evidence in, Plaintiffs' Evidence Out

Of course, the current practice in civil and criminal cases both supports and undermines the claim that novel neuroscience may go the way of the polygraph. On the supporting side, courts already seem to show intense skepticism toward cutting-edge neuroscientific techniques and have generally excluded such evidence.<sup>78</sup> Moreover, because such evidence is offered almost always by plaintiffs in support of recovery for claims against defendants, exclusion is consistent with courts' historical skepticism of plaintiff-proffered novel scientific proof.<sup>79</sup> Thus, the general and specific patterns point in consistent directions: toward exclusion of the evidence.

But undermining the probability of an enduring ban are the cases from the criminal context that already dispute that prediction. Although courts have generally excluded novel neuroscience, recall that careful inspection reveals one significant exception: neuroscientific proof offered by defendants in sentencing proceedings.<sup>80</sup> That suggests both a willingness to embrace some role in service of the criminal defendants and not the prosecution.

\*633 Current observations thus only partly conform to the general pattern of novel scientific evidence--plaintiffs still remain largely rebuffed, but defendants can find some favor with courts. But the story is not yet fully told, because courts, for the most part, have yet to confront the question of admitting novel neuroscience when offered by prosecutors.<sup>81</sup> It thus may still unfold that the customary patterns prove enduring; courts generally exclude plaintiffs' novel neuroscience applying strict admissibility tests, while admitting prosecutors' evidence under a more relaxed standard. In this case, the only surprise would be that defendants will also benefit from such evidence when proffered for mitigation purposes.

This kind of modified status quo is not that unimaginable, as described in greater depth below.<sup>82</sup> In fact, it is this familiar story that causes many to fear that neuroscience represents a "double-edged sword"--what appears on its face a boon for criminal defendants, who can claim "my brain did it," will in fact be a weapon for prosecutors, who will use neuroscientific findings to argue for the incorrigibility, remorselessness, antisocial tendencies, or deviance of defendants.

## 3. The Status Quo, Revised

A third possibility, however, is that the current trend holds even as prosecutors seek to marshal neuroscientific evidence in support of their claims. Courts would extend the general skepticism shown to plaintiffs who offer novel techniques to prosecutors, even while continuing to carve out a role for the criminal defendant. It is not quite a ban because criminal defendants are permitted limited use. And again, because prosecutors have yet to offer such evidence with regularity, this scenario constitutes pure conjecture. But it may be that the heyday of admission of thinly supported \*634 scientific evidence is over, and the kind of rigorous attention given to plaintiffs' evidence will now be given to prosecutors' evidence as well. The raised awareness of the problem of wrongful conviction, and the prominent role that faulty science has played in those injustices,<sup>83</sup> could contribute to a sense that courts ought to shore up their admissibility standards when it comes to novel scientific evidence offered by the government in a criminal case. Recent admonitions against admitting flawed forensics may also cause courts to examine such evidence with greater intensity.<sup>84</sup>

But if that is the case, and courts roundly reject novel neuroscience, then how could current trends permitting defense introduction of such evidence hold and not collapse into the total ban scenario? There is little specific law on the operation of scientific admissibility standards as applied to criminal defendants as opposed to the prosecution, but what little exists suggests that there is no meaningful difference. While there is some legal support for the notion that a defendant's constitutional rights to confrontation and due process may override even rules intended to safeguard evidentiary reliability,<sup>85</sup> that line of reasoning has long lain fallow. If so, then *Daubert* should demand as much from criminal defendants as it does from prosecutors, and much novel neuroscience would be excluded.

But why might courts not back away from admission when it comes to defense mitigation? Three reasons.

First, the mantra that "death is different" is now so familiar that it practically needs no citation.<sup>86</sup> The Supreme Court consistently has distinguished capital cases in its review of the constitutionality of sentences, applying a much more robust concept of Eighth Amendment proportionality and even the Sixth Amendment right to counsel than evident in noncapital cases. Thus, it would be consistent with this disparity to also admit defense neuroscientific evidence that would not have passed muster if offered by the prosecution or civil parties.

Second, this tacit recognition that the Constitution applies differently to death cases finds explicit expression in the law of evidentiary admissibility for capital mitigation hearings. The Constitution requires that juries be allowed to consider "any relevant mitigating factor"<sup>87</sup> offered in a capital sentencing hearing, and "virtually no limits are placed on the relevant mitigating evidence a capital defendant may introduce concerning his own \*635 circumstances."<sup>88</sup> It is also clear that ordinary rules of evidence--which typically do not apply in full form during sentencing proceedings in any event<sup>89</sup>--must yield in capital mitigation hearings.<sup>90</sup> At the same time, there remains debate as to whether and to what extent *Daubert*, or a similar reliability-based standard for expert evidence, applies in the sentencing context.<sup>91</sup>

Finally, to the extent that the debate centers on introduction of such evidence in capital mitigation hearings--as opposed to the sentencing phase for noncapital offenses as well--then in practicality it will be an issue in only the handful of states that continue to impose the widely rejected sentence of capital punishment.<sup>92</sup>

In short, it is possible to imagine, and compatible with a commitment to the consistent application of legal standards, that novel neuroscientific evidence becomes a regular feature of capital mitigation hearings, even if rejected for every other proffered use. Plaintiffs, prosecutors, and other litigants (including defendants seeking to use novel neuroscience as proof in the liability phase) may continue to meet resistance from courts skeptical that \*636 such evidence can meet the threshold showing of reliability, even as criminal defendants in mitigation hearings make full use of such evidence.

#### 4. Final Thoughts About Spillover Effects

One final scenario requires elaboration. Although there are good reasons, founded both in law and legal practice, to expect that novel neuroscience will initially remain largely cabined to capital mitigation and other serious sentencing hearings, it is easy to imagine that mounting pressure would result in its adoption in other contexts. If, in fact, the routine use of novel neuroscientific evidence in mitigation hearings were to result in such pressure to apply elsewhere, what might that expansion look like? To what other proceedings might it most naturally reach?

Already, novel neuroscience has had an impact outside of the capital sentencing context: namely, in the noncapital sentencing context, albeit in a categorical and nonindividualized way. In *Graham v. Florida*<sup>93</sup> and *Miller v. Alabama*,<sup>94</sup> the Supreme Court relied heavily on neuroscientific studies to limit the reach of life without parole sentences for juvenile offenders, based on studies showing the immaturity of their brains. But apart from continuing to rely on neuroscience in this categorical fashion-- isolating categories for addicted persons or the mentally ill, for instance-- the real breakthrough would be to apply neuroscientific findings to noncapital, individual sentencing determinations. Indeed, *Miller* opened the door precisely to that kind of evidence. By holding that courts cannot impose mandatory life without parole, but must make individualized

determinations in the case of juveniles,<sup>95</sup> the Court opened the door to the consideration of individual neuroscientific findings in support of a particular defendant's claim. And if a juvenile can use brain development as a mitigating factor, why not a twenty-something-year-old?

Also, there are already slight signs of prosecutors' interest in using neuroscientific evidence,<sup>96</sup> and it is easy to imagine, as many have, the myriad ways in which prosecutors might make further use of it in the future. It is easiest to imagine such uses in contexts that, like capital sentencing, do not suffer from the constraint of strict (or clear) evidentiary rules, such as bail hearings, competency determinations, and noncapital sentencing. Still, other proceedings, such as civil commitment hearings predicated on mental illness or future dangerousness, may require adherence to *Daubert* and *Frye* but not require the stringent burden of proof that must be met for criminal proceedings.

\*637 Moreover, as noted earlier, judges who become used to seeing neuroscientific proof in capital sentencing hearings may believe, as a matter of basic fairness, that the prosecution ought to be permitted to respond in kind with its own evidence. For example, in Professor Farahany's study of criminal cases, she noted that prosecutors do not always respond solely with argument to defense efforts to use neuroscientific testing--they sometimes use the neuroscientific proof itself to argue against the defendant.<sup>97</sup> In such cases, "[s]ome of the brain abnormality evidence introduced by a criminal defendant at trial can cut against him at a civil commitment hearing,"<sup>98</sup> as happened in the case of a man who had suffered a serious brain injury that purportedly explained his aggression but whose injury prosecutors used to also show his incapacity for reform.<sup>99</sup>

A kind of "good for the goose, good for the gander" reasoning also appears in existing sentencing law. In *Payne v. Tennessee*,<sup>100</sup> as noted earlier, the Supreme Court expressly stated that "virtually no limits are placed on the relevant mitigating evidence a capital defendant may introduce concerning his own circumstances."<sup>101</sup> But the Court also added that "[t]he State has a legitimate interest in counteracting the mitigating evidence which the defendant is entitled to put in."<sup>102</sup> Otherwise, there is "the potential for such unfairness."<sup>103</sup> Once the defendant introduces exculpatory neuroscientific proof, it seems only natural that courts would allow prosecutors to respond in kind. And when such evidence takes the form of novel neuroscience, courts may prove reluctant to reject prosecutorial evidence as insufficiently reliable having admitted the same kind of evidence when offered by the defense. In this way, evidence that now serves the interests of defendants, propelled to admission by a defendant's special constitutional rights,<sup>104</sup> may quickly become precedent relied upon by courts to admit the same kind of evidence more broadly, even when offered against the defense. If so, government use of neuroscientific proof could be grandfathered in through defense standards that were never that onerous, resulting in a new kind of *Daubert* equilibrium.

Of course, as courts grow accustomed to hearing neuroscientific evidence in bail proceedings, sentencing proceedings, competency determinations, and the like, will they remember that such evidence did not have to meet stringent hurdles of reliability when confronted with neuroscientific evidentiary proffers during the guilt phase of a trial? Current case reviews suggest that \*638 courts generally reject such evidence, even when offered by the defense, although instances of admission occur.<sup>105</sup> But will that pattern endure even if the science does not meaningfully evolve? Will it not seem odd to a judge to rule evidence unreliable that, in a hearing months earlier, the judge cited as part of a basis for a detention decision? Similarly, it is easy to imagine that opinions admitting such evidence at the request of defendants citing constitutional values will be successfully cited by prosecutors seeking to introduce the same kind of evidence on their own behalf.

If novel neuroscience gains a foothold in the parts of a criminal case that are not characterized by extensive discovery, robust adversarial hearing, or formal evidence rules (including *Daubert*- or *Frye*-type reliability screens), then it will no doubt have an advantage when it first starts cropping up in the more demanding phases of adjudication. Indeed, the lamentable state of public defense in the United States suggests that many unfounded neuroscientific claims may go altogether unchallenged even if there were legitimate legal and scientific bases to keep such evidence out.<sup>106</sup>

Finally, might this embrace of novel neuroscientific evidence, once a regular feature of criminal cases, eventually bleed over to the civil context as well? Will an opinion that admits evidence of "toxic encephalopathy"<sup>107</sup> in mitigation become a supportive citation for a motion to admit such evidence when offered by a civil plaintiff? Although the traditional narrative about the divide between civil and criminal *Daubert* suggests that distinctions between the two can be maintained, it is not



inconceivable that the inroads made in the criminal context might ease the path for at least some civil plaintiffs going forward. The same judges that hear neuroscientific evidence in capital cases inevitably will preside over other kinds of proceedings, including civil matters.

As such evidence becomes increasingly familiar and judges acclimate to its particular vernacular, the novelty of using neuroscientific proof may start to wear off and earlier boundaries dissolve. After all, judges tend to expect a baseline of reliability from all evidence. Even in a capital hearing, most judges would not allow the defendant to present an astrologer who would testify that the defendant only acted because Mercury was in retrograde. Judges may feel a fundamental discomfort with the idea of a discrepancy--the notion that neuroscience is somehow reliable enough for a death sentence determination but not for less serious offenses or monetary claims.

Furthermore, empirical evidence shows that neuroscientific proof is susceptible to motivated reasoning, “the unconscious tendency to assimilate \*639 information in a manner biased towards reaching a particular outcome.”<sup>108</sup> Thus, judges may begin to deem such evidence reliable when it confirms other proof, or even their own intrinsic beliefs about a particular condition, and incline toward a more generous *Daubert* or *Frye* standard in noncapital or civil cases.

## CONCLUSION

Novel neuroscientific evidence now stands at the precipice of the judicial system, seeking entry. But that system’s history of safeguarding scientific proof suggests that even if neuroscience is ready, the courts may not be. On the civil side, courts have struggled to show evenhandedness and consistency in the degree to which they subject plaintiffs’ evidence to scrutiny, often being accused of reserving their most intense oversight for plaintiffs’ proffers. On the criminal side, courts have repeatedly shown themselves willing to allow the most spurious forms of “science” when offered by prosecutors--with catastrophic consequences.

The disconnect between these two realities--courts’ leniency toward criminal prosecutors and harshness toward civil plaintiffs--has henceforth created little overt tension in our appraisal of the rule of law because the scientific methods proffered in each category varied markedly. Apart from fire science and handwriting analysis, which have their own unique pathologies, the scientific techniques rejected by civil courts had little bearing on the methods prosecutors sought to introduce.

Novel neuroscience, however, may stand alone at the crossroads of civil and criminal evidence. Like lie detection, its closest analogue, novel neuroscience offers something of value to both civil and criminal litigants and to plaintiffs, prosecutors, and defendants. As courts confront questions of its admissibility, then, they will have to squarely confront the demonstrated problem of inconsistent application of admissibility standards. What will be the result? Only time will tell. But whatever the outcome, observers may gain a clearer sense of the successes and failures of our evidentiary admissibility standard.

## Footnotes

<sup>a1</sup> Professor, New York University School of Law. I am grateful to Professor Deborah Denno and the participants of the *Fordham Law Review*’s symposium entitled *Criminal Behavior and the Brain: When Law and Neuroscience Collide*, held at Fordham University School of Law, for their helpful feedback and inspiring comments in connection with this Article. I owe thanks to Ayelet Evrony and Peter Varlan, who provided superb research assistance, as well as to the Filomen D’Agostino and Max E. Greenberg Research Fund, which supported this work. For an overview of the symposium, see Deborah W. Denno, *Foreword: Criminal Behavior and the Brain: When Law and Neuroscience Collide*, 85 FORDHAM L. REV. 399 (2016).

<sup>1</sup> See 5 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY §§ 20:3-:17 (2015); Deborah W. Denno, *The Myth of the Double-Edged Sword: An Empirical Study of Neuroscience Evidence in Criminal Cases*, 56 B.C. L. REV. 493, 499 (2015); see also Nita A. Farahany, *Neuroscience and*

*Behavioral Genetics in US Criminal Law: An Empirical Analysis*, 2 J.L. & BIOSCIENCES 485, 491 (2016) (disputing that neuroscientific evidence is used exclusively and largely unsuccessfully in capital mitigation hearings, but reporting one study that shows that it is largely used, with success, in that context).

<sup>2</sup> Farahany, *supra* note 1, at 2 n.3.

<sup>3</sup> *Id.* at 10.

<sup>4</sup> 509 U.S. 579 (1993).

<sup>5</sup> See, e.g., Leslie Morsek, Comment, *Get on Board for the Ride of Your Life! The Ups, the Downs, the Twists, and the Turns of the Applicability of the "Gatekeeper" Function to Scientific and Non-Scientific Expert Evidence: Kumho's Expansion of Daubert*, 34 AKRON L. REV. 689, 704 (2001).

<sup>6</sup> 293 F. 1013 (D.C. Cir. 1923).

<sup>7</sup> 522 U.S. 136 (1997).

<sup>8</sup> 526 U.S. 137 (1999).

<sup>9</sup> The iconic text that influenced this view was written by Peter W. Huber. See PETER W. HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM 3 (1991) (warning that "any self-styled scientist, no matter how strange or iconoclastic his views, will be welcome to testify in court").

<sup>10</sup> See, e.g., David S. Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685, 690, 703-09 (2000) (critiquing the opinion's "limited framework" for understanding the scientific enterprise).

<sup>11</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1315 (9th Cir. 1995).

<sup>12</sup> *Id.* at 1317 n.5.

<sup>13</sup> Note, *Admitting Doubt: A New Standard for Scientific Evidence*, 123 HARV. L. REV. 2021, 2024 & nn.24-25 (2010) (collecting sources on opposing sides of the debate).

<sup>14</sup> See, e.g., Andrew W. Jurs & Scott DeVito, *Et Tu, Plaintiffs?: An Empirical Analysis on Daubert's Effect on Plaintiffs, and Why Gatekeeping Standards Matter (a Lot)*, 66 ARK. L. REV. 975, 984-85 & n.47 (2013) (canvassing studies estimating high rates of expert evidence, in reporting on findings about plaintiff and defendant filing and removal behavior in light of the evidentiary standard).

<sup>15</sup> NAT'L RESEARCH COUNCIL, NAT'L ACADS., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 11 (2009).

NEUROSCIENCE AND THE CIVIL/CRIMINAL DAUBERT..., 85 Fordham L. Rev. 619

---

- 16 David L. Faigman, *Admissibility Regimes: The "Opinion Rule" and Other Oddities and Exceptions to Scientific Evidence, the Scientific Revolution, and Common Sense*, 36 SW. U. L. REV. 699, 716 (2008).
- 17 *See, e.g., supra* note 9; *infra* notes 24-25.
- 18 D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99, 99 (2000).
- 19 *Id.* at 143.
- 20 *Id.* at 102-04.
- 21 *Id.* at 143-49, app.
- 22 *Id.* at 143-49.
- 23 *See generally* NAT'L RESEARCH COUNCIL, *supra* note 15, at 11; Paul C. Giannelli, *Daubert and Criminal Prosecutions*, 26 CRIM. JUST. 3 (2011).
- 24 A. Leah Vickers, *Daubert, Critique and Interpretation: What Empirical Studies Tell Us About the Application of Daubert*, 40 U.S.F. L. REV. 109, 109-10, 126-37 & nn.3-4 (2000) (reviewing and citing studies); *see also* Margaret A. Berger, *What Has a Decade of Daubert Wrought?*, 95 AM. J. PUB. HEALTH S59, S64 (2005) (citing studies that show that "judges are much more likely since *Daubert* to scrutinize expert testimony before trial and then to limit or exclude expert testimony" in civil cases, although "courts are not applying *Daubert* stringently in the criminal context"); Peter J. Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggestions for Reform*, 95 AM. J. PUB. HEALTH S107, S109 (2005). *But see* Andrew Jurs, *Gatekeeper with a Gavel: A Survey Evaluating Judicial Management of Challenges to Expert Reliability and Their Relationship to Summary Judgment*, 83 MISS. L.J. 325, 335-38 (2014) (reviewing prior studies and noting greater degrees of uncertainty).
- 25 Jennifer L. Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 PSYCHOL. PUB. POL'Y & L. 339, 364 (2002).
- 26 *See generally* Paul C. Giannelli, *Forensic Science: Under the Microscope*, 34 OHIO N.U. L. REV. 315 (2008) (surveying disciplines like bite marks, handwriting, and ballistics).
- 27 NAT'L RESEARCH COUNCIL, *supra* note 15, at 7-8.
- 28 *See, e.g., Julie A. Seaman, A Tale of Two Dauberts*, 47 GA. L. REV. 889, 892 n.12 (2013).
- 29 Jane Campbell Moriarty, *Will History Be Servitude?: The NAS Report on Forensic Science and the Role of the Judiciary*, 2010 UTAH L. REV. 299.
- 30 *Id.* at 315 (footnote omitted).



- 31 See H.R. 988, 104th Cong. (1995). Georgia, however, still maintains some distinction. See Seaman, *supra* note 28, at 891 & n.9 (citing a provision that reads: “In criminal proceedings, the opinions of experts on any question of science, skill, trade, or like questions shall always be admissible; and such opinions may be given on the facts as proved by other witnesses” (quoting GA. CODE ANN. § 24-7-707 (2011))).
- 32 Samuel R. Gross, *Expert Evidence*, 1991 WISC. L. REV. 1113, 1119 (reporting on findings from a study of California courts).
- 33 See, e.g., NAT’L RESEARCH COUNCIL, *supra* note 15, at 36 (noting that forensic analysis typically aims for identification, individualization, association and reconstruction).
- 34 Gross, *supra* note 32, at 1119.
- 35 Jennifer Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725, 759-60, 774 (2011) (discussing the lack of training of analysts and institutional relationship between forensic laboratories and police departments).
- 36 See ERIN E. MURPHY, *INSIDE THE CELL: THE DARK SIDE OF FORENSIC DNA* 4-5 (2015) (listing differences between medical and forensic DNA testing).
- 37 See *id.* at 4.
- 38 See Seaman, *supra* note 28.
- 39 *Id.* at 898.
- 40 *Id.* at 897-99.
- 41 *Id.* at 907-08.
- 42 Of course, there is also the question of whether fire science is more rigorous than handwriting analysis and thus deserving of admission. Certainly some aspects of fire investigation can lay substantially more claim to scientific legitimacy than handwriting analysis. However, fire science unfortunately has been marked by a longstanding history of experts whose conclusions rest more on myth and folklore than on scientific truth. See generally Caitlin M. Plummer & Imran J. Syed, “*Shifted Science*” Revisited: *Percolation Delays and the Persistence of Wrongful Convictions Based on Outdated Science*, 64 CLEV. ST. L. REV. 483, 485-95 (2016).
- 43 Seaman, *supra* note 28, at 904, 907.
- 44 In Professor Seaman’s sample, in all but one of the criminal cases the evidence was offered by the prosecution, and in civil cases nearly all were offered by the plaintiff. *Id.* at 899.
- 45 See *id.* at 901, 908.

46 *Id.* at 900.

47 *See, e.g.*, Owen D. Jones et al., *Neuroscientists in Court*, 14 NATURE 730, 730 (2013) (“In ... a steadily increasing number of similar cases in both criminal and civil courts, neuroscientific evidence has been introduced to support a party’s legal claim as well as to argue its irrelevance or invalidity (by the opposing party).”).

48 Stephen J. Morse, *Introduction*, in A PRIMER ON CRIMINAL LAW AND NEUROSCIENCE, at xv (Stephen J. Morse & Adina L. Roskies eds., 2013).

49 *See, e.g.*, David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417 (2014).

50 *See, e.g.*, OWEN JONES, JEFFREY SCHALL & FRANCIS SHEN, LAW AND NEUROSCIENCE 28 (2014); A. Philip Dawid, David L. Faigman & Stephen E. Fienberg, *Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects?*, 43 SOC. METHODS & RES. 359 (2014).

51 JONES, SCHALL & SHEN, *supra* note 50, at 269-302.

52 *See generally* *Graham v. Florida*, 560 U.S. 48 (2010); *Roper v. Simmons*, 543 U.S. 551 (2005).

53 *See generally* 5 DAVID L. FAIGMAN ET AL., *supra* note 1, §§ 20:1-:63.

54 Two major surveys both reached the same conclusion. *See* Denno, *supra* note 1, at 493 (“My analysis reveals that neuroscience evidence is usually offered to mitigate punishments in the way that traditional criminal law has always allowed, especially in the penalty phase of death penalty trials.”); Farahany, *supra* note 1, at 7 (conceding that neurobiological evidence is used most often in criminal cases for mitigation purposes, but claiming that it also is gaining ground in other areas, such as competency determinations or capacity defenses). Professor Farahany’s findings of usage beyond just mitigation hinge in large part on a more capacious definition of “what ‘counts’ ... as neurobiological evidence,” because her study includes any reference to medical history of brain trauma or interviews aimed at determining such history and not just scans or physical evidence of brain injury. *Id.* at 10.

55 *See* 5 DAVID L. FAIGMAN ET AL., *supra* note 1, § 20:11.

56 *See id.* §§ 20:12-:13; *see also* Farahany, *supra* note 1, at 19.

57 *See* 5 DAVID L. FAIGMAN ET AL., *supra* note 1, § 20:10.

58 *See* Farahany, *supra* note 1, at 21.

59 *See* 5 DAVID L. FAIGMAN ET AL., *supra* note 1, § 20:4.

60 *See id.* §§ 20:5-:6. QEEG and SPECT, short for quantitative electroencephalogram and single-photon emission computerized

tomography, respectively, are neuroimaging methods. *Id.*

61 *See id.* § 20:7.

62 *See id.* §§ 20:8,:16.

63 Jones et al., *supra* note 47, at 730.

64 *See* 5 DAVID L. FAIGMAN ET AL., *supra* note 1, §§ 20:5-:8.

65 *See, e.g.*, JONES ET AL., *supra* note 50, at 731 (describing worries about judicial misuse of neuroscientific evidence).

66 *See supra* Part II.A.

67 NAT'L RESEARCH COUNCIL, NAT'L ACADS., THE POLYGRAPH AND LIE DETECTION 18 (2003) ("In popular culture and media, the polygraph device is often represented as a magic mind-reading machine. These facts reflect the widespread mystique or belief that the polygraph test is a highly valid technique for detecting deception ....").

68 The sense that polygraphs seem to favor criminal defendants is evident in the U.S. Attorneys' Manual, which instructs prosecutors to seek to exclude polygraph evidence. 5 DEP'T OF JUSTICE MANUAL tit. 9, no. 262 (3d ed. 2015) (outlining arguments for attorneys on exclusion).

69 *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

70 *See, e.g.*, Charles W. Daniels, *Using Polygraph Evidence After Scheffer* (pt. 2), CHAMPION, June 2003, at 36.

71 Timothy B. Henseler, Comment, *A Critical Look at the Admissibility of Polygraph Evidence in the Wake of Daubert: The Lie Detector Fails the Test*, 46 CATH. U. L. REV. 1247, 1248 (1997).

72 *See id.* at 1250 ("[T]he potential for admissibility of polygraph evidence appears to be greater than ever before.").

73 *See* NAT'L RESEARCH COUNCIL, *supra* note 67, at 203 ("In fact, topics such as bite mark and hair identification, fingerprinting, arson investigation, and tool mark analysis have a less extensive record of research on accuracy than does polygraph testing.").

74 *See* Henseler, *supra* note 71, at 1278-79; *see also* *United States v. Scheffer*, 523 U.S. 303 (1998) (holding that the bar on polygraph evidence did not violate defendant's constitutional rights); Robin D. Barovick, Comment, *Between Rock and a Hard Place: Polygraph Prejudice Persists After Scheffer*, 47 BUFF. L. REV. 1533 (1999).

75 *See* 5 DAVID L. FAIGMAN ET AL., *supra* note 1, § 38:5 (discussing "evaluations of polygraph evidence").

76 *See supra* Part II.A.



- 77 See 5 DAVID L. FAIGMAN ET AL., *supra* note 1, § 38:1 n.2 (“[T]here are a multitude of other techniques and technologies heralded as the next generation of lie detector. Principal among these competitors might be the use of functional magnetic resonance imaging (fMRI) for this purpose.”).
- 78 See *supra* Part II.A.
- 79 See *supra* notes 16-18, 24 and accompanying text.
- 80 See *supra* note 54.
- 81 Professor Denno’s study concludes that prosecutors presently use such evidence “only rarely.” Denno, *supra* note 1, at 499. Professor Farahany is more equivocal, noting that “[p]rosecutors, too, have seized on cognitive neuroscience to argue that defendants are incorrigible and should be given longer sentences” and “to denigrate defendants’ characters and to demonstrate defendants’ likely future dangerousness.” Farahany, *supra* note 1, at 4-5. But it is unclear whether this refers to arguments made by prosecutors in response to evidence offered by the defense and intended as mitigating (e.g., the touted “double-edged sword” of neuroscience, in which the defense argues “my brain is defective, spare me” while the prosecution counters that “defendant’s brain is defective, incarcerate him”), as opposed to those marshaled to support neuroscientific evidence offered ab initio by the prosecutor. See, e.g., *id.* at 21 (recounting prosecutors’ argument to this effect in response to defense evidence). In Professor Farahany’s article surveying existing cases, she expounds the facts of cases that seem to consist exclusively of defense-offered evidence. See *id.* at 12, 14-19 (discussing competency challenges raised by defense regarding standing trial, tendering a plea, and confessing; support for mental illness or mens rea defenses; involuntariness; and sentencing). Notably, Professor Denno found that when “prosecutors did utilize neuroscience evidence to suggest a defendant’s propensity to commit crimes, they typically did so only by building upon the evidence first introduced by a defense expert.” Denno, *supra* note 1, at 526. She further found that only eighty cases contained future dangerousness discussions grounded in neuroscience, and only in ten of those was that “neuroscience evidence introduced by the defense ... leveraged by the prosecution in an effort to prove the defendant’s future dangerousness.” *Id.* at 528.
- 82 See *infra* Part II.B.4.
- 83 See generally BRANDON L. GARRETT, CONVICTING THE INNOCENT: WHERE CRIMINAL PROSECUTIONS GO WRONG 84-117 (2012).
- 84 See, e.g., NAT’L RESEARCH COUNCIL, *supra* note 15, at 96-110.
- 85 See, e.g., *Rock v. Arkansas*, 483 U.S. 44 (1987); *Chambers v. Mississippi*, 410 U.S. 284 (1973).
- 86 Nonetheless, Rachel Barkow’s excellent review of the “two tracks” of sentencing law provides a helpful primer. Rachel E. Barkow, *The Court of Life and Death: The Two Tracks of Constitutional Sentencing Law and the Case for Uniformity*, 107 MICH. L. REV. 1145, 1147 (2009) (noting courts’ insistence that “death is different” and arguing that it is both legally unsupported and theoretically unwise).
- 87 *Eddings v. Oklahoma*, 455 U.S. 104, 112 (1982) (citing *Lockett v. Ohio*, 438 U.S. 586 (1978)).
- 88 *Payne v. Tennessee*, 501 U.S. 808, 822 (1991). Note, however, that the Court has expressly stated that “relevant” has no special

meaning in the capital context. *Tennard v. Dretke*, 542 U.S. 274, 284 (2004) (noting that the evidence must be that which “tends logically to prove or disprove some fact or circumstance”).

<sup>89</sup> See, e.g., 18 U.S.C. § 3593(c) (2012) (stating that “[i]nformation is admissible regardless of its admissibility under the rules governing admission of evidence at criminal trials except that information may be excluded if its probative value is outweighed by the danger of creating unfair prejudice, confusing the issues, or misleading the jury”); FED. R. EVID. 1101(d)(3). Interestingly, there have been constitutional challenges to this rule that reveal how it operates both to the benefit and detriment of each side. For instance, it may help criminal defendants by permitting less robustly reliable evidence in mitigation, but most defense advocates deem the standard as harming defendants because it lessens the bar for the reliability of aggravation evidence offered by the prosecution in support of aggravation. Challenges along both lines have largely failed. See, e.g., *United States v. Snarr*, 704 F.3d 368, 399 (5th Cir. 2013) (reviewing the Federal Death Penalty Act’s “relaxed evidentiary standard” and reaffirming its constitutionality); see also Michael D. Pepson & John N. Sharifi, *Two Wrongs Don’t Make a Right: Federal Death Eligibility Determinations and Judicial Trifurcations*, 43 AKRON L. REV. 1, 13 (2010). At the same time, defendants have mounted *Daubert* and *Frye* challenges to evidence introduced at sentencing, such as to “scientific findings” claiming future dangerousness, most often without much success. See *infra* note 90.

<sup>90</sup> See *Green v. Georgia*, 442 U.S. 95, 97 (1979) (holding that the hearsay rule could not serve to exclude testimony during the capital penalty phase); cf. *Oregon v. Guzek*, 546 U.S. 517, 525 (2006) (holding that the Constitution did not grant the right to the defendant to introduce innocence-related alibi evidence that undermined conviction during the penalty phase).

<sup>91</sup> Some courts have expressly held that *Daubert* does not apply, see, e.g., *United States v. Fields*, 483 F.3d 313, 342 (5th Cir. 2007), while others have simply skirted the issue, see, e.g., *United States v. Barnette*, 211 F.3d 803, 815 (4th Cir. 2000) (rejecting the defense’s *Daubert*-based challenge to “[p]sychopathy checklist” evidence at sentencing without resolving the applicability of *Daubert*); *Smithers v. State*, 826 So.2d 916 (Fla. 2002) (noting conflicting evidence on PET scan without referencing admissibility standards).

<sup>92</sup> Although nineteen states formally retain the death penalty, only seven states have executed an individual in the past two years. See DEATH PENALTY INFO. CTR., FACTS ABOUT THE DEATH PENALTY, <http://www.deathpenaltyinfo.org/documents/FactSheet.pdf> (last updated Aug. 30, 2016) [<https://perma.cc/UFQ4-C4YS>].

<sup>93</sup> 560 U.S. 48 (2010). As noted earlier, the Court also held it unconstitutional to execute juveniles in *Roper v. Simmons*, 543 U.S. 551 (2005). See *supra* note 52 and accompanying text.

<sup>94</sup> 132 S. Ct. 2455 (2012).

<sup>95</sup> *Id.* at 2475.

<sup>96</sup> See, e.g., Denno, *supra* note 1, at 526 (arguing that the concern is overblown that prosecutors will use neuroscientific evidence to bolster arguments of future dangerousness); Farahany, *supra* note 1, at 12-17 (discussing competency).

<sup>97</sup> See Farahany, *supra* note 1, at 22.

<sup>98</sup> *Id.*

<sup>99</sup> See *id.*

100 501 U.S. 808 (1991).

101 *Id.* at 822.

102 *Id.* at 825 (quoting *Booth v. Maryland*, 482 U.S. 496, 517 (1987) (White, J., dissenting)).

103 *Id.*

104 *See, e.g.*, *United States v. Sandoval-Mendoza*, 472 F.3d 645, 656 (9th Cir. 2006) (reversing exclusion of brain tumor evidence that bolstered entrapment defense); *State v. Ferrell*, 277 S.W.3d 372, 381 (Tenn. 2009) (reversing exclusion of “toxic encephalopathy” evidence that supported the defendant’s claim that he was too cognitively impaired to have plotted escape).

105 *See, e.g.*, 5 DAVID L. FAIGMAN ET AL., *supra* note 1, §§ 20:9-:16.

106 *See, e.g.*, Risinger, *supra* note 18, at 135 (“When I first started looking at these post*Daubert* cases, I expected to find records of multiple well-litigated attacks on the weakest kinds of common prosecution-proffered expertise, with any system bias coming from judicial decisions. What I found was an apparent systematic failure to seriously litigate these issues on the part of the criminal defense bar.”).

107 *See Ferrell*, 277 S.W.3d at 375, 381.

108 Nicholas Scurich & Adam Shniderman, *The Selective Allure of Neuroscientific Explanations*, PLOS ONE 2 (Sept. 2014), <http://journals.plos.org/plosone/article/asset?id=10.1371/journal.pone.0107529.PDF> [<https://perma.cc/W8AX-T76P>].

85 FDMLR 619

---

End of Document

© 2018 Thomson Reuters. No claim to original U.S. Government Works.



92 N.Y.U. L. Rev. 1532

New York University Law Review  
November, 2017

Note  
Stephanie L. Damon-Moore<sup>1</sup>

Copyright © 2017 by Stephanie L. Damon-Moore

## TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM

*In the last decade, many fields within forensic science have been discredited by scientists, judges, legal commentators, and even the FBI. Many different factors have been cited as the cause of forensic science's unreliability. Commentators have gestured toward forensic science's unique development as an investigative tool, cited the structural incentives created when laboratories are either literally or functionally an arm of the district attorney's office, accused prosecutors of being overzealous, and attributed the problem to criminal defense attorneys' lack of funding, organization, or access to forensic experts.*

*But none of these arguments explain why trial judges, who have an independent obligation to screen expert testimony presented in their courts, would routinely admit evidence devoid of scientific integrity. The project of this Note is to understand why judges, who effectively screen evidence proffered by criminal defendants and civil parties, fail to uphold their gatekeeping obligation when it comes to prosecutors' forensic evidence, and how judges can overcome the obstacles in the path to keeping bad forensic evidence out of court.*

INTRODUCTION	1533
<sup>1</sup> THE FORENSIC SCIENCE PROBLEM	1537
<sup>II</sup> PROSECUTORS, DEFENDERS, AND JUDGES: WHO CAN FIX THE FORENSIC SCIENCE PROBLEM <sup>2</sup>	1541
<sup>A</sup> Prosecutors	1541
<sup>B</sup> Criminal Defense Attorneys	1546
<sup>C</sup> Judges	1550

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

III UNDERSTANDING THE JUDICIAL FAILURE TO GATEKEEP	1555
<sup>A</sup> <i>Understanding How Judges Overlook Problems with Forensic Science</i>	1555
<sup>1</sup> <i>Judges Lack Scientific Knowledge</i>	1555
<sup>2</sup> <i>Judges Do Not Receive Helpful Information from Defense Attorneys</i>	1558
<sup>3</sup> <i>Judges Rely on Heuristics to Admit Prosecution Evidence</i>	1559
<sup>B</sup> <i>From Recognition to Action</i>	1563
<sup>1</sup> <i>Why Judges May Choose Not to Exclude Bad Evidence</i>	1563
<sup>2</sup> <i>The Path to Solving the Forensic Science Problem</i>	1566
CONCLUSION	1570

**\*1533 INTRODUCTION**

In the last decade, many fields within forensic science have been discredited by scientists,<sup>1</sup> judges,<sup>2</sup> legal commentators,<sup>3</sup> and even the FBI.<sup>4</sup> The foundation, methodology, execution, and conclusions of forensic scientists have been repeatedly called into question.<sup>5</sup> Criticisms have been directed at all forensic disciplines, from bite-mark **\*1534** analysis<sup>6</sup> to DNA analysis.<sup>7</sup> Hundreds if not thousands of convictions that rested in whole or in part on this “junk science” have been overturned,<sup>8</sup> while innumerable additional convictions have been cast into doubt. Yet this evidence has been and continues to be routinely admitted in criminal court.<sup>9</sup>

Commentators have pointed to a number of possible explanations for the problems with forensic science evidence: forensic science’s unique development as an investigative tool;<sup>10</sup> the structural incentives created when laboratories are either literally or functionally an arm of the district attorney’s office;<sup>11</sup> and inadequate education programs for forensics.<sup>12</sup> They have blamed the proliferation of **\*1535** invalid forensic science testimony in the courtroom on overzealous prosecutors<sup>13</sup> and criminal defense attorneys’ incompetence, lack of funding, and lack of access to forensic experts.<sup>14</sup>

But none of these explanations account for why trial judges, who have an independent obligation to screen expert testimony presented in their courts, routinely admit evidence and permit testimony devoid of scientific integrity. And given the rigor with which trial judges screen experts proffered by criminal defendants and civil litigants, no serious argument can be made that judges lack the scientific savvy to execute their duty responsibly.<sup>15</sup> Nor, for that matter, are the shortcomings of forensic science especially complex.<sup>16</sup> Yet forensic evidence has been treated as reliable, even when the experts themselves acknowledge glaring gaps in their scientific methods and conclusions.<sup>17</sup>

Although scholars have noted that trial judges do not rigorously uphold their expert admissibility screening obligation when it comes to prosecutors’ forensic experts, few have tried to explain why.<sup>18</sup> One **\*1536** notable exception is retired Judge Donald Shelton of Michigan’s 22nd Circuit Court in Ann Arbor, Michigan, currently the director of the Criminal Justice Studies Program at the University of Michigan-Dearborn.<sup>19</sup> Judge Shelton has argued that judicial permissiveness with regard to forensic evidence is most reasonably attributed to judicial bias--or “attitudinal blinders”--and has suggested some of the same causes that I will discuss, including the fact that many judges are former prosecutors, as well as the influence of “tough on crime” elections.<sup>20</sup> Although I agree with Judge Shelton that judicial bias likely influences expert admissibility decisions,<sup>21</sup> my analysis recognizes a number of constraints on judges’ abilities to recognize and address problems with forensic science, including judges’ lack of scientific training,<sup>22</sup> the frequent absence of defense objections to prosecution proffers,<sup>23</sup> and

concerns--rooted in both law and policy--about withholding relevant evidence from juries.<sup>24</sup>

For the purposes of this Note, I define the “forensic science problem” as the admission of unreliable forensic testimony in criminal trials. My framing reflects the view that the worst outcome of “junk science” is wrongful convictions,<sup>25</sup> and accordingly that preventing \*1537 wrongful convictions is the highest priority for reform. Others may define the worst outcome as the failure to convict the correct perpetrator, allowing a guilty person to remain “on the streets.” This definition, in turn, may prompt a focus on solutions such as increased federal funding for forensic science, or improvements to forensic science education programs. My definition, of course, points to a different kind of solution: keeping bad evidence out.

This Note proceeds as follows: Part I summarizes a few of the most common--and problematic--shortcomings of forensic science evidence. Part II examines the relative abilities and incentives of prosecutors, defense attorneys, and judges to solve the forensic science problem, and concludes that only judges are well positioned to do so. Lastly, Part III addresses the big question: How can judges uphold their gatekeeping responsibilities and keep junk science out of the courtroom?

## I THE FORENSIC SCIENCE PROBLEM

In recent years, nearly all forensic disciplines have been criticized as insufficiently tested or even invalid.<sup>26</sup> Though a comprehensive analysis of the problems with various forensic disciplines is beyond the scope of this Note,<sup>27</sup> it is worth fleshing out some of these issues, to both bring the forensic science problem to life and to dispel the notion that there is something special about forensic evidence that makes it especially difficult to screen out. A brief foray into the character of the problems with forensic evidence makes it clear that the admission of flawed forensic science cannot be so easily explained.

The paradigmatic forensic science evidence involves a comparison of two samples--one taken from a known suspect, the other found at the scene of the crime--and the conclusion that there is a “match.”<sup>28</sup> Whether through the statements of the witness or the \*1538 surmise of the jury, this match is interpreted to demonstrate with virtual certainty that the defendant was at the scene of the crime. The validity of this conclusion rests on two interlocking assumptions: first, that a match is defined by some kind of industry standard, as opposed to the personal view of the testifying expert, and second, that this match is inculpatory because the particular combination of features that matched the exemplar is either literally unique or at least highly unusual.

Research, however, has shown that these necessary assumptions are not present in many forensic fields. Several forensic disciplines have been criticized for their lack of an industry-wide match standard. For example, bite-mark analysis, also called forensic odontology, has no industry-wide standard defining how many points of similarity must exist to call a comparison a “match.” Instead, the reference manual for the American Board of Forensic Odontology’s “Human Bitemark Analysis Guidelines” is comprised of a cursory, bullet-point-style list of features to be documented, along with suggestions as to how the bite mark could be compared to exemplars.<sup>29</sup> Although the guidelines rather unhelpfully list some vocabulary for articulating whether a bite mark does or does not match an exemplar, notably absent are any *standards* for determining whether said match exists.<sup>30</sup> In other words, bite-mark “experts” have some parameters for the types of criteria to consider and the way to articulate their findings, but the critical middle step--drawing the correct conclusions from the evidence--is left to individual discretion.

Remarkably, forensic odontologists have not been shy about the fact that their “scientific” field is entirely devoid of standards for matching samples.<sup>31</sup> For example, in a leading case from the Missouri Court of Appeals, *State v. Sager*, the bite-mark expert acknowledged that there was “no standard procedure for arriving at conclusions” in the field of forensic odontology, and that his methodology “would not \*1539 necessarily be used by all experts.”<sup>32</sup> Along the same lines, Mississippi’s Supreme Court has repeatedly recognized that there “is little consensus in the scientific community on the number of points which must match before any positive identification can be announced” and that “methods of comparison employed in a particular case may differ.”<sup>33</sup>

Yet, in these and other cases, the evidence gets admitted. In *Sager*, for example, the court not only allowed the expert to testify that the bite mark “matched,” but also allowed him to testify that “no two people could have identical mouths,” that “the bite mark reflected in the photograph of the breast of the victim was *beyond a reasonable doubt* placed upon the victim’s breast by appellant,” and that “the conclusions reached [by all experts] would be the same.”<sup>34</sup> This confidence is misplaced. In the limited scientific studies of bitemark analysis, forensic odontologists disagreed not only about whether there was a match to the exemplar, but even about whether the mark was made by a human.<sup>35</sup>

Bite-mark analysis may be the poster child for a forensic field that commentators love to hate,<sup>36</sup> but another forensic discipline that lacks standards for a “match” is a former darling of the forensic evidence world: fingerprint analysis.<sup>37</sup> The FBI once proudly wrote that “[o]f all the methods of identification, fingerprinting alone has proved to be both infallible and feasible.”<sup>38</sup> But while courts have been \*1540 admitting fingerprint evidence for over a hundred years,<sup>39</sup> the discipline lacked an industry-wide match standard until the National Research Council criticized this shortcoming in a 2009 report.<sup>40</sup> Since then, the Scientific Working Group on Friction Ridge Analysis, Study and Technology has issued standards that, according to the President’s Council of Advisors on Science and Technology (PCAST), “begin[ ] to move latent print analysis in the direction of an objective framework.”<sup>41</sup> Nevertheless, PCAST concluded that fingerprint analysis has “a considerable way to go” before achieving objectivity.<sup>42</sup>

If and when a match *can* be reliably established in any forensic discipline, the significance of the match hinges on the degree to which a match is *unusual*. Consider, for example, blood types: A serologist might find that the defendant’s blood type matches a sample found at the crime scene. How inculpatory that is depends precisely on how unusual the blood type is. For instance, eleven percent of the white population have Type B blood.<sup>43</sup> So, if you have a white Type B defendant, and you have a Type B sample taken from the scene, you can safely say that the sample excludes eighty-nine percent of the white population and does not exclude the defendant.

On the other hand, if you don’t know anything about how common each blood type is, you could testify that you found a match, but not about how unusual that match is. Take, for example, microscopic hair analysis. We don’t know how common it is for people within a racial group to have similar hair characteristics, which means that testimony about hair characteristics ought to be limited to a statement that the sample found at the scene “could have” come from the defendant.<sup>44</sup> Unfortunately, microscopic hair experts have struggled to restrain themselves: For decades, experts testified that only about “5 percent of the population” shared certain characteristics, or that finding a match put the odds of a false positive at “one chance in 10,000.”<sup>45</sup> In 2015, an FBI review concluded that, of the twenty-eight \*1541 FBI agents who conducted microscopic hair analyses, *twenty-six* made erroneous statements in written reports or oral testimony.<sup>46</sup>

This brief description of a few common problems with forensic evidence demonstrates that the shortcomings of forensic science are not so complex that prosecutors cannot understand them, defense attorneys cannot challenge them, and judges cannot bar evidence because of them. Yet over and over again forensic expert witnesses are permitted to give invalid testimony. The following Part looks at the three primary actors within the criminal justice system--prosecutors, criminal defense attorneys, and trial judges--to determine who is best suited to address the forensic science problem.

## II PROSECUTORS, DEFENDERS, AND JUDGES: WHO CAN FIX THE FORENSIC SCIENCE PROBLEM?

### A. Prosecutors

Prosecutors have a great degree of power to prevent erroneous admissions. As the actors who proffer the evidence in the first place, they could all but solve the forensic science problem tomorrow simply by declining to present evidence that lacks a valid foundation and a reliable methodology.<sup>47</sup> But, despite the expectation that prosecutors will act as “minister[s] of justice,”<sup>48</sup> the adversarial structure of the criminal justice system incentivizes zealous prosecutions.<sup>49</sup> Scholars have repeatedly noted the tension between the stated neutrality of the prosecutor and the reality of prosecution in the United States.<sup>50</sup> While \*1542 a prosecutor’s role is ostensibly to reach the truth, some prosecutors have evinced marked resistance to the truth when



the truth happens to be that the defendant is innocent.<sup>51</sup>

States' experts are not immune from the conviction mindset, either, perhaps because of the ways in which forensic laboratories are aligned with law enforcement. For example, many forensic science disciplines were initially developed by law enforcement for investigative purposes.<sup>52</sup> And in most jurisdictions, forensic laboratories are still literally a part of the prosecutors' office.<sup>53</sup> Even where there is no formal connection, forensic labs get the vast majority of their business from prosecutors and law enforcement, potentially creating a sense \*1543 that success on the job means finding a match and getting a conviction.<sup>54</sup>

One example of the dangers of such a "team" mentality among prosecutors and experts comes from the area of Shaken Baby Syndrome (SBS). Despite being based on what can charitably be described as a thin medical foundation,<sup>55</sup> growing national awareness of child abuse, among other factors, led to an explosion in prosecutions for SBS in the United States in the nineties and the first decade of the twenty-first century.<sup>56</sup> Many of these prosecutions came out of child abuse units in hospitals and collaboration between prosecutors and child abuse physicians in cases where a caregiver was alleged to have shaken an infant, causing brain damage or death.<sup>57</sup> In her seminal book on SBS prosecutions, Professor Deborah Tuerkheimer describes the role physicians played in SBS cases: "Doctors came to court and explained that, notwithstanding the absence of any other signs of abuse, shaking could be proved by three neurological symptoms: bleeding beneath the outer layer of membranes surrounding the brain, bleeding in the retina, and brain swelling."<sup>58</sup> These doctors would go on to testify that the shaking must have been the cause, that it must have been unreasonably violent shaking, and that the perpetrator must have been the last person with the infant.<sup>59</sup> The medical testimony thus wasn't merely the most damning part of the case; it *was* the case.

Against that backdrop, the medical community writ large began to call these conclusions into question. In 2011, for example, an advisory board member of the National Center on Shaken Baby Syndrome described the three symptoms as a "myth," dismissing the notion that any "trained pediatrician" could equate them with abuse.<sup>60</sup> But child abuse units full of physicians trained to make just \*1544 such a conclusion were already in place. A study by Northwestern University that examined 3000 SBS cases from a twenty-five-year period exposed what is arguably the result of collaboration between prosecutors and physicians: SBS prosecution "hot spots."<sup>61</sup> For example, Sarpy County, Nebraska, and nearby Douglas County, Nebraska, are number one and number four respectively in SBS prosecutions per capita.<sup>62</sup> These blockbuster numbers of prosecutions may reflect the local "team" strategy, in which "law enforcement, child advocacy centers, prosecutors, the Department of Health and Human Services, and medical professionals" work together to investigate possible child abuse.<sup>63</sup> As this example demonstrates, adversarial incentives influence not only prosecutors themselves, but may influence experts with whom they work closely as well.

Not surprisingly given the adversarial structure of the criminal justice system, prosecutors have shown little interest in policing their own evidence. The National District Attorneys Association (NDAA), for instance, responded to the publication of the 2016 PCAST Report<sup>64</sup> by criticizing the committee's methods and composition and rejecting all of its findings. "Experience shows these disciplines offer reliable and powerful evidence in a court of law,"<sup>65</sup> the NDAA argued. The NDAA continued: "It is therefore entirely inappropriate for the report to suggest otherwise to this country's courts."<sup>66</sup> Attorney General Loretta Lynch was no more interested in adopting PCAST's findings, despite being appointed by the same liberal administration that commissioned the report. Her response was that the Department of Justice "remain[s] confident that, when used properly, forensic science evidence helps juries identify the guilty and clear the innocent," and "[w]hile we appreciate [PCAST's] contribution to the field of scientific inquiry, the department will not be adopting the recommendations \*1545 related to the admissibility of forensic science evidence."<sup>67</sup>

More recently, President Trump's appointment for Attorney General, Jeff Sessions, took further steps to halt forensic science reform efforts. For instance, Sessions recently announced he would halt a collaboration on forensic science research between leading scientists and the Department of Justice (DOJ).<sup>68</sup> In 2013, responding to the NRC Report, the Obama administration formed the National Commission on Forensic Science (NCFS), a panel of leading scientists charged with advising the executive branch on efforts to standardize and validate forensic disciplines.<sup>69</sup> But in April of 2017, Sessions announced that the NCFS charter would not be renewed, over the objection of several of NCFS's scientists.<sup>70</sup> "For too long," the panel members wrote, "decisions regarding forensic science have been made without the input of the research science

community.”<sup>71</sup> This exclusion, they argued, led to a “disconnection between the fundamental principles of science and some forensic disciplines.”<sup>72</sup>

In addition to allowing the NCFS charter to expire, Sessions also announced that he would be terminating a review of FBI testimony in several forensic disciplines. Reviews in recent years had already turned up widespread problems with the testimony of FBI analysts in the area of microscopic hair analysis, and Deputy Attorney General Sally Yates had announced that expanded review would focus on identifying “whether the same kind of ‘testimonial overstatement’ that we found during our review of microscopic hair evidence could have crept into other disciplines.”<sup>73</sup> This review has now been “suspended” \*1546 by the Trump administration.<sup>74</sup> As Peter Neufeld, co-founder of the Innocence Project, said of these developments, “the [DOJ] has literally decided to suspend the search for the truth.”<sup>75</sup>

As these actions on the part of the DOJ demonstrate, prosecutors’ influence extends beyond the courtroom and into policy and funding decisions. The fact that Sessions could unilaterally disband the NCFS, for example, is a product of the DOJ’s efforts under Lynch to maintain control of forensic science reform. It would therefore be shortsighted to blame the Trump administration for the forensic science problem, since in the months following publication of the PCAST Report in September of Obama’s final year in office, “not only [did] the Obama administration [do] nothing about the report, the Justice Department ... publicly denounced it.”<sup>76</sup> As long as forensic science continues to be “powerful evidence in a court of law,” as the NDAA put it,<sup>77</sup> the DOJ is likely to exercise that influence to keep reform efforts within their control. And if courts give the benefit of the doubt to prosecutors as to the reliability of their evidence, prosecutors will have no incentive to fund research that might prove otherwise.

In sum, neither the structural incentives for prosecutors nor their actual practices suggest that prosecutors are likely to solve the forensic science problem.

### ***B. Criminal Defense Attorneys***

Criminal defense attorneys, unlike prosecutors, are structurally incentivized to challenge forensic testimony, yet commentators have noted that they rarely do so.<sup>78</sup> There are a few possible explanations for this phenomenon. First, because most criminal defendants are indigent, most criminal defense attorneys are public defenders or panel attorneys.<sup>79</sup> Average caseload numbers for public defenders far \*1547 exceed recommended numbers,<sup>80</sup> forcing them to cut corners and file only the most critical motions, which could explain the relative quiet when it comes to motions regarding forensic science evidence.<sup>81</sup> Panel attorneys are seldom compensated at competitive rates for their public appointments,<sup>82</sup> incentivizing them to spend as little time as possible on their indigent clients relative to their paying ones. They also often lack experience with criminal trials.<sup>83</sup> And given the frequency with which forensic evidence is admitted,<sup>84</sup> even defense attorneys with time and money to burn might be reluctant to occupy the court’s time with a motion challenging a prosecutor’s proffer, fearing that doing so will annoy trial judges who have been admitting such evidence for years. That said, criminal defense attorneys do challenge experts sometimes, and will likely do so more often in light of the growing chorus of criticisms from the NRC Report, the PCAST \*1548 Report, and other commentators. Their motions, however, are routinely denied.<sup>85</sup>

Still, when bad forensic evidence is admitted--whether in the absence of a motion to exclude, or in the face of one--defense attorneys can challenge this evidence with the usual tricks of the trade: rigorous cross-examination and dueling experts.<sup>86</sup> But neither of these traditional safeguards is sufficient. Cross-examination, for example, comes only after the jury has watched the judge--ostensibly the only “neutral” lawyer in the room--qualify the expert at the outset of his or her testimony.<sup>87</sup> Jurors likely attribute at least a modicum of reliability to someone whose credentials have just been read and used as the basis for admittance.<sup>88</sup> Effective cross-examination also requires the defense attorney to be relatively knowledgeable about the shortcomings of the evidence, which may be too tall an order for overwhelmed public defenders or appointed counsel more accustomed to writing wills or appearing in bankruptcy court.<sup>89</sup> A study of 137 transcripts from the trials of DNA exonerees exposed routine failure on the part of defense attorneys to effectively cross-examine expert witnesses, even when their testimonies were patently false.<sup>90</sup> And finally, even a great cross-examination may not be effective against a confident expert witness, since cross-examination is best suited to exposing personal flaws like “[v]eracity, memory, motivation, prejudices,

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

---

and \*1549 biases,” rather than scientific invalidity.<sup>91</sup> Cross-examination is therefore both rare and likely ineffective as a tool to prevent invalid forensic science from swaying a jury toward conviction.

In light of these shortcomings of cross-examination, one could argue that contrary experts proffered by the defense are the best safeguard to prevent the jury from crediting the invalid testimony of a prosecution expert. Defense attorneys experience problems in this arena as well, however. First, they may not have access to qualified experts, either for lack of funding<sup>92</sup> or because the majority of experts are employed by the state.<sup>93</sup> Judges may also admit the state’s expert and then turn around and bar a similarly credentialed expert from the defense: At least one study has shown that judges are significantly more likely to exclude experts proffered by the defense than those proffered by the state.<sup>94</sup> And even if they are admitted, juries faced with conflicting expert testimonies may discount the testimonies of *both* experts, perhaps interpreting them as “hired guns” who are simply testifying as directed.<sup>95</sup> Thus the traditional safeguards of the trial process are often too little, too late.

**\*1550 C. Judges**

Although prosecutors lack structural incentives to police forensic evidence appropriately, and defense attorneys lack the resources and authority to do so, judges do not face either of these obstacles. As the neutral arbiter of the courtroom, the judge is tasked with ensuring the fairness and integrity of criminal proceedings. This nebulous expectation is concretized in the trial judge’s affirmative obligation to screen all expert evidence before it is presented at trial, known as the judicial gatekeeping function.

All fifty states and the federal government are bound by rules of evidence that require judges to screen expert evidence before it is presented to a jury.<sup>96</sup> For example, Federal Rule of Evidence 702 states that “[a] witness who is qualified as an expert ... may testify in the form of an opinion or otherwise” only if certain criteria are met, including that “the testimony is based on sufficient facts or data,” that “the testimony is the product of reliable principles and methods,” and that those principles and methods were “reliably applied” in the case at bar.<sup>97</sup> Although there’s no question that judges benefit from rigorous challenges to evidence by criminal defense attorneys,<sup>98</sup> the fact remains that with or without a defense motion, judges are obligated to protect the jury from certain expert evidence.<sup>99</sup>

The majority of states and the federal government apply the admissibility test established by the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,<sup>100</sup> known as the *Daubert* test.<sup>101</sup> *Daubert* superseded the longstanding test articulated in 1923 by the Court of Appeals of the District of Columbia in *Frye v. United States*.<sup>102</sup> The *Frye* standard asks simply whether the methods or principles in question “have gained general acceptance in the field in \*1551 which it belongs,” known as the “general acceptance” test.<sup>103</sup> In *Daubert*, the Court interpreted Rule 702 to mandate a judicial finding of “evidentiary reliability” for admittance.<sup>104</sup> The *Daubert* opinion provides four factors for courts to consider in determining the reliability of expert evidence, including (1) whether the theory or technique at issue can be tested, (2) whether it has been subject to peer review and publication, (3) the known rate of error, and (4) whether the theory or technique is generally accepted in the relevant scientific community.<sup>105</sup> The Court emphasized, however, that the test is “flexible”; these factors are neither mandatory nor are they necessarily exhaustive considerations.<sup>106</sup> The Court has subsequently written that the *Daubert* factors “do *not* constitute a ‘definitive checklist or test’”<sup>107</sup> and their applicability “depend[s] on the nature of the issue, the expert’s particular expertise, and the subject of his testimony.”<sup>108</sup>

Because *Daubert* interprets Rule 702, which applies identically to civil and criminal cases, the case requires judges to apply the same “exact[ing] standard[ ]”<sup>109</sup> to expert proffers from any party. But the flexibility of the *Daubert* test makes it ripe for biased decisionmaking. Over the course of three cases, known as the “*Daubert* trilogy,”<sup>110</sup> the Supreme Court granted trial judges discretion on three levels. First, it is up to the judge to decide what kind of procedure to follow when making an admissibility determination—for instance, whether to hold \*1552 a *Daubert* hearing as opposed to deciding the issue on paper.<sup>111</sup> Second, which, if any, *Daubert* factors to consider is reserved to the judge’s discretion.<sup>112</sup> And third, the ultimate decision of whether to admit the evidence is a matter of discretion.<sup>113</sup> Recognizing the multiple layers of deference developed through the *Daubert* doctrine, Justice Scalia concurred in one of the trilogy cases with the sole purpose of cautioning that the discretion that trial judges enjoy “is not discretion to perform the function inadequately” but is only “discretion to choose

among *reasonable* means of excluding expertise that is *fausse* and science that is junky.”<sup>114</sup>

In the aftermath of the **Daubert** trilogy, scholars predicted a sea change in the admission of forensic science.<sup>115</sup> But within a few years, those predictions were replaced with observations that only prosecutors’ experts seemed to be evading review. A famous (or infamous) pair of rulings from the Eastern District of Pennsylvania in *United States v. Llera Plaza*<sup>116</sup> illustrates both how **Daubert** can be applied to bar shoddy forensic evidence, and how forensic evidence makes it into court despite its shortcomings.

In *Llera Plaza I*, Judge Louis Pollack applied the relatively new **Daubert** standard to fingerprint evidence proffered by the government, and found it lacking in all four factors. First, as to whether the technique “can be (and has been) tested,”<sup>117</sup> Judge Pollack found that the methodology used by the FBI had not been subject to relevant testing, rejecting the government’s contention that “testing” included \*1553 “adversarial’ courtroom testing.”<sup>118</sup> Second, he considered whether fingerprinting had been subject to peer review, and concluded that because the only “peer review” came from the analysts themselves, who “tend to be skilled professionals who have learned their craft on the job and without concomitant advanced academic training,”<sup>119</sup> the FBI’s methodology had not been subject to “peer review” in the scientific sense.<sup>120</sup> Third, Judge Pollack found evidence regarding the error rate of fingerprint identification “unpersuasive, one way or another” and found that fingerprint comparison lacked “uniformly accepted ‘scientific’ standards.”<sup>121</sup> Fourth, he found that because the field of fingerprint analysis is not itself a reliable field, general acceptance within the field of fingerprint examiners “by itself cannot sustain the government’s burden in making the case for admissibility of fingerprint testimony under [Rule] 702.”<sup>122</sup> In sum, fingerprint analysis had not been tested, was not standardized, could not support claims of a low error rate, and was not generally accepted within a disinterested scientific community.<sup>123</sup>

Nevertheless, after finding fingerprint analysis lacking at every turn, Judge Pollack ruled that fingerprint examiners *could* testify, albeit cabining them to “descriptive, not judgmental” testimony.<sup>124</sup> In so doing, he alluded to a consideration not sanctioned by the **Daubert** Court, namely the “century of judicial acquiescence in fingerprint identification” that would render total exclusion “unwarrantably heavy-handed.”<sup>125</sup>

Then, the plot thickened. Under protestations from the government that cabining expert testimony would “seriously compromise[ ]” the government’s “prosecutorial effectiveness,” Judge Pollack held another **Daubert** hearing and overturned his original ruling.<sup>126</sup> In *Llera Plaza II*, the judge did not dispute the defense’s contention that no new factual or legal basis had emerged to justify the reconsideration, but simply lowered the bar for the FBI. For instance, instead of hewing to his original judgment that the government had not persuaded \*1554 him that the error rate was acceptably low,<sup>127</sup> Judge Pollack now held that “there is no evidence that the error rate of certified FBI fingerprint examiners is unacceptably high.”<sup>128</sup> This way of looking at the error rate is all the more remarkable because under Rule 702 the burden is clearly on the government to show that the evidence is reliable by a preponderance of the evidence.<sup>129</sup> Flipping this standard on its head, Judge Pollack concluded that he had not been “persuaded that courts should *defer* admission of testimony with respect to fingerprinting.”<sup>130</sup> In short, under pressure from the DOJ and the FBI, Judge Pollack put the burden on the defense to show *unreliability* rather than holding the prosecution to its burden of proving *reliability*.

The layers of discretion baked into **Daubert** open the door for this kind of special treatment. For instance, in the Ninth Circuit’s **Daubert** opinion following remand from the Supreme Court, Chief Judge Kozinski offered additional factors for trial judges to consider in making **Daubert** rulings, and called it “very significant” whether the expert’s testimony “grow[s] naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying.”<sup>131</sup> And yet, with no explanation for the distinction, he dropped a footnote exempting “law enforcement” evidence from this consideration:

There are, of course, exceptions. Fingerprint analysis, voice recognition, DNA fingerprinting and a variety of other scientific endeavors closely tied to law enforcement may indeed have the courtroom as a principal theatre of operations. As to such disciplines, the fact that the expert has developed an expertise principally for purposes of litigation will *obviously* not be a substantial consideration.<sup>132</sup>



Interestingly, Judge Kozinski has now joined the crusade against “voodoo science” proffered by prosecutors.<sup>133</sup> But his 1995 opinion \*1555 reflects the evidently popular sentiment that law enforcement evidence is somehow uniquely exempt from the rigorous screening that other evidence must pass through.

Prosecutors, criminal defense attorneys, and trial judges are all complicit to some degree in the admission of shoddy forensic science evidence. Only trial judges, however, are explicitly tasked with keeping unreliable expert evidence out of the courtroom. The existence of this affirmative obligation, however, has not prevented unreliable evidence proffered by prosecution from being admitted in court and functioning as the basis for criminal convictions. The following Part attempts to make sense of this reality.

### III UNDERSTANDING THE JUDICIAL FAILURE TO GATEKEEP

As the preceding discussion shows, the forensic science problem is unlikely to be solved by either prosecutors or defense attorneys. That would seem to leave us with trial judges, and indeed Rule 702 and *Daubert* suggest that judges have the means to stop bad forensic evidence in its tracks. But judges and other commentators have suggested a number of reasons why judges do not step in to keep bad forensic expert evidence out of court. This section considers those arguments, and concludes that despite the obstacles in their paths, judges can and should take a leadership role in solving the forensic science problem.

#### *A. Understanding How Judges Overlook Problems with Forensic Science*

##### *1. Judges Lack Scientific Knowledge*

One commonly offered explanation for why judges are not well suited to the task of solving the forensic science problem is that they lack scientific training. Indeed, some evidence suggests that judges are not well equipped to operationalize *Daubert* because, though fully capable of reciting the *Daubert* factors, they are not well versed in what those factors actually mean for the evidence in question.<sup>134</sup>

A common problem in the forensic science context, for example, is defining “peer review” for the purposes of *Daubert*’s second \*1556 factor.<sup>135</sup> In *Daubert* the Court described peer review as “submission to the scrutiny of the scientific community,” which “increases the likelihood that substantive flaws in methodology will be detected.”<sup>136</sup> One of the most problematic aspects of forensic science is that, because virtually all forensic disciplines (with the notable exception of DNA analysis) were developed for law enforcement purposes, there is no neutral “scientific community” reviewing the methodologies developed in the field. Although some judges have recognized that review by practitioners working in law enforcement cannot reasonably be equated with the kind of dispassionate review described in *Daubert*,<sup>137</sup> others accept arguments from prosecutors and practitioners that publications and reviews by fellow practitioners satisfy the peer review consideration.<sup>138</sup>

The *Llera Plaza* opinions, analyzed above,<sup>139</sup> illustrate one type of interpretive error with regard to the peer review factor. In *Llera Plaza I*, Judge Pollack found that fingerprint analysis came up short on peer review because, to the extent that publications were submitted to a peer community, that community was comprised of “skilled professionals who have learned their craft on the job and without any concomitant advanced academic training” and thus was not a “scientific community.”<sup>140</sup> In *Llera Plaza II*, however, Judge Pollack reversed course, this time equating the field of fingerprint analysts to “accountants, vocational experts, accident-reconstruction experts, appraisers of land or of art, experts in tire failure analysis, or others” who “have ‘technical, or other specialized knowledge’” and need not represent a “scientific community” in order to satisfy the “peer review” factor.<sup>141</sup> In fact, in both instances Judge Pollack’s analysis overlooks a critical component of peer review, which is that peer review is designed in part to control for bias, and cannot serve that \*1557 function adequately when the entire peer group shares a systemic bias. Because fingerprint analysis was developed by law enforcement for law

enforcement, the entire peer community may share a systemic bias toward conviction and ought to be approached with skepticism. Having interpreted the Supreme Court's analysis in *Daubert* to mean simply that the peer community ought to be comprised of *highly educated* experts, rather than *neutral* experts, it is no surprise that Judge Pollack backtracked from his initial ruling when confronted with the numerous fields in which scientific training is irrelevant.

But while judges may generally be ill equipped to rigorously evaluate expert evidence proffered by any party, this lack of scientific expertise may unfairly advantage prosecutors and disadvantage defendants. For example, a survey of federal cases in the seven years following the *Daubert* decision found that when criminal defense attorneys challenged prosecution experts, the government prevailed in sixty-one out of the sixty-seven appellate opinions.<sup>142</sup> When the prosecutor was challenging the ruling, on the other hand, the defendant lost forty-four out of fifty-four cases, while seven of the remaining ten cases were remanded for a *Daubert* hearing.<sup>143</sup> Similar patterns have been identified elsewhere: A subsequent analysis of nearly seven hundred state and federal judicial opinions published in the five-and-a-half years after *Daubert* showed an enormous disparity in success between prosecution and defense proffers of expert evidence at the trial court level, with prosecution experts being admitted 95.8% of the time while defense experts were admitted a mere 7.8% of the time.<sup>144</sup>

Prosecution evidence also appears to be held to a lower standard than evidence in civil cases. Empirical studies following *Daubert* have shown that "evidence proffered by plaintiffs in civil cases receives harsh scrutiny for reliability, whereas evidence proffered by prosecutors \*1558 in criminal cases typically gets a free pass."<sup>145</sup> Nor is it the case that evidence proffered in civil cases is somehow easier for people without scientific training to understand. As Professor Jane Moriarty wryly puts it, "[i]n civil cases, courts seem quite up to the task of evaluating microbiology, teratology, and toxicology evidence," but "when it comes to evaluating the shortcomings of lip prints and handwriting, courts are unable to muster the most minimal grasp of why a standardless form of comparison might lack evidentiary reliability."<sup>146</sup>

## 2. Judges Do Not Receive Helpful Information from Defense Attorneys

Judges faced with prosecution expert proffers may not receive the information they need to make a fair gatekeeping determination from the usual source: the opposing party.<sup>147</sup> Ordinarily, judges rely heavily on the arguments and evidence that the parties present to them.<sup>148</sup> As analyzed above, however, defense attorneys frequently fail to challenge prosecution experts.<sup>149</sup> Without those challenges, judges make admissibility determinations solely on the basis of the prosecutor's expert notice and the judge's knowledge and beliefs about forensic science.<sup>150</sup> And even where challenges are made, they may be made \*1559 without the input--and accompanying affidavit-- of a contrary expert witness, and may as a result be substantively weak or lacking in credibility (particularly in contrast to the proffered prosecution expert).<sup>151</sup> A judge could reasonably interpret a defense objection unsupported by an expert's advice and affidavit as a Hail Mary motivated by vigorous advocacy rather than a well-founded challenge to invalid evidence.

Still, with or without a credible defense challenge, judges have a responsibility to keep problematic expert evidence out of the courtroom. As difficult as this may be without the help of effective defense challenges, a growing body of literature--including the NRC Report<sup>152</sup> and the PCAST Report,<sup>153</sup> as well as scholarship,<sup>154</sup> journalism,<sup>155</sup> and even case law<sup>156</sup>--is available to educate judges as they make these decisions. Thus while judges may lack information from the usual source--the opposing party--the information itself is available, and judges have a responsibility to find it.

## 3. Judges Rely on Heuristics to Admit Prosecution Evidence

Judges faced with complex scientific questions for which they have little or no training, often combined with an absence of adequate information from the defense, may fail to grasp the shortcomings of \*1560 forensic evidence because they are relying on heuristics, or mental shortcuts, to sidestep the substantive question altogether. Humans often rely on heuristics when faced with complex problems, either because they fail to realize that the shortcut is not an appropriate proxy for the more difficult decision, or because they lack the motivation or energy needed to engage in the effortful work demanded.<sup>157</sup>

Perhaps the single best example of an inappropriate but tempting heuristic for judges to rely on in the forensic science context is how long certain types of forensic evidence, such as fingerprint analysis, have been admitted as evidence.<sup>158</sup> As one district judge put it, “[T]he methods of latent print identification can be and have been tested. They have been tested for roughly 100 years. They have been tested in adversarial proceedings with the highest possible stakes--liberty and sometimes life.”<sup>159</sup> The problem with this reasoning is that a mistake repeated for a hundred years is still a mistake.<sup>160</sup> Professor Jane Moriarty has illustrated this fallacy by comparing the long-history argument in the forensic science context to similar arguments made to justify medical use of leeches, a practice that survived for thousands of years until it was subjected to scientific testing.<sup>161</sup> A long history of admission, in other words, is a poor proxy for validity or reliability of a methodology.

Judges may also rely on their instincts about the parties to guide their decisionmaking when it comes to forensic science experts.<sup>162</sup> Heuristics in this line are particularly concerning because they are likely to militate in favor of the prosecution and against the defendant overall. One reason why such an imbalance may exist is that many \*1561 judges were prosecutors before they donned their robes. For example, a review of the judges nominated during President Obama’s first seven years in office concluded that while he made significant strides toward diversifying the bench in terms of gender, race, and sexual orientation, the professional experience of his nominees was strikingly homogenous, with the largest number of nominees coming from private practice, followed by prosecutors’ offices.<sup>163</sup> In state courts, the trend is the same. For example, one 2009 study of state supreme court justices found that 32.9% of state supreme court justices had prosecutorial experience, while only 15.4% had experience as public defenders.<sup>164</sup> A high representation of prosecutorial experience on the bench may both reflect and contribute to a judicial preference for prosecutors. And, interlocking with a bias in favor of consistency as described above, former prosecutors-turned-judges may be biased in favor of the evidence itself, as they likely proffered similar forensic evidence during their time as prosecutors--and got convictions because of it.

Furthermore, in some jurisdictions the court system is organized such that prosecutors are assigned to specific courtrooms. Judges and prosecutors assigned to these courtrooms may overlap on a daily basis, leading to a collegial, team-like atmosphere.<sup>165</sup> In the context of heavily discretionary decisionmaking like evidentiary rulings, a friendly relationship with a prosecutor may unconsciously guide a judge’s decision to decline a motion for a **Daubert** hearing, discourage her from weighing factors that cut against admission, and ultimately push her toward allowing invalid testimony.

In addition to potential bias in favor of prosecutors, judges may also harbor unconscious biases against criminal defendants. Such biases could come from a number of sources. One factor worth considering is that criminal defendants are disproportionately poor and Black,<sup>166</sup> in a society in which implicit bias against disadvantaged class \*1562 and racial groups “has proven to be extremely widespread.”<sup>167</sup> Given the prevalence of these biases, the judiciary would be in a class by itself if it were unaffected by them.<sup>168</sup> That is not to say that simply because “judges are human”<sup>169</sup> there are no differences in the degree to which they harbor or act on bias. To the contrary, empirical evidence suggests that judges are *not* as susceptible to all cognitive biases as is the population as a whole.<sup>170</sup> That said, studies also indicate that judges are susceptible to cognitive biases, including racial bias.<sup>171</sup> It would therefore be too simple to conclude either that judges as a whole are indistinguishable from the general populace, or that, because their jobs demand neutrality, they can simply will away the influence of any spurious considerations.

But although judges may be tempted to rely on heuristics like whether the evidence has a long history of being admitted to court, or instincts about the parties themselves, this is not an insurmountable problem. As the authors of one study of judicial decisionmaking suggest, judges can reduce error and the influence of systemic bias in part by simply making admissibility decisions slowly and deliberately.<sup>172</sup> Similarly, just recognizing that **Daubert** determinations are vulnerable \*1563 to heuristic-based decisionmaking may combat the influence of such bias.

### ***B. From Recognition to Action***

What the foregoing analysis boils down to is this: Although judges may understandably be hampered in their efforts to recognize the flaws of the forensic evidence proffered by prosecutors, the obstacles to gatekeeping are not so great or so insurmountable as to justify admission of untested or junk evidence. But for judges who recognize the forensic science

problem and are prepared to keep it out of court, the final problem may be simply that these judges feel powerless to break with history, particularly where the result is keeping evidence away from the jury.

### *1. Why Judges May Choose Not to Exclude Bad Evidence*

As a threshold matter, it is worth noting that in both state and federal courts, judges are generally encouraged to take a liberal stance with regard to admitting evidence. The baseline requirement for evidence to be admissible is merely that it be “relevant,” meaning that it “has any tendency to make a fact more or less probable than it would be without the evidence,” and that the fact in question “is of consequence in determining the action.”<sup>173</sup> Emphasizing that this is not a high bar, the comment to Rule 401 advises that “a brick is not a wall,” meaning that evidence need not prove an entire case on its own, it need only be helpful.<sup>174</sup> Relevance, in other words, is a low standard, and forensic evidence easily satisfies it.<sup>175</sup>

In addition to the guidance provided by the Federal Rules of Evidence, principles stemming from sources as lofty as the Constitution may encourage a liberal mindset when it comes to the admission of evidence. For example, Judge Jack Weinstein of the Eastern District of New York has provided the following explanation for a liberal approach to evidence rulings:

The jury’s constitutionally based fact-finding primacy demands a measure of forbearance on the part of judges. We cannot forget that, because of our narrow life experiences, our ability to draw appropriate inferences from the evidence in the cases before us is \*1564 limited. Whenever it is arguably appropriate, we should allow the matter to go to the jury, reserving the right to set aside its decision if there proves to be no rational basis for the verdict. Not only is this the fairest approach in most cases, but it also provides litigants with something most desire--a chance to be heard and a judgment by their fairly selected peers.<sup>176</sup>

In sum, judges making expert admissibility determinations are encouraged by everything from the rules of evidence to the broadest principles of justice to err on the side of admitting evidence.

In the expert context, there are countervailing principles to consider. As Judge Patti Saris of the District of Massachusetts has opined, “The Court’s vigilant exercise of [the Rule 702] gatekeeper role is critical because of the latitude given to expert witnesses to express their opinions ... and because an expert’s testimony may be given greater weight by the jury due to the expert’s background and approach.”<sup>177</sup> Yet even where experts are concerned, the Court has implied that trial judges should take a liberal approach to admission, emphasizing that “[v]igorous 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.”<sup>178</sup> A judge who recognizes weaknesses in forensic evidence that is routinely admitted may therefore feel compelled to admit it despite its flaws.

Ironically, the ultimate safeguard against judicial error--appellate review--may actually *discourage* judges from gatekeeping effectively. In at least one jurisdiction, the law is clear that once the admission of a “new scientific technique” has been approved by an appellate court in a written opinion, “the precedent so established may control subsequent trials.”<sup>179</sup> Although many jurisdictions, including federal courts, have not adopted this approach,<sup>180</sup> judges in those jurisdictions are no doubt wary of departing from regular practice. \*1565 Former Judge Nancy Gertner of the District of Massachusetts, an outspoken skeptic of forensic evidence,<sup>181</sup> alludes to this concern in *United States v. Green*.<sup>182</sup> “Although the scholarly literature is extraordinarily critical [of ballistics evidence],” she writes, “court after court has continued to allow the admission of this testimony.”<sup>183</sup> Judge Gertner concludes that precedent unanimously militated in favor of admitting the evidence, even while she described this type of reasoning as “troubling.”<sup>184</sup> “It runs the risk of ‘grandfathering in irrationality,’” she cautions, “without reexamining it in the light of *Kumho* and *Daubert*.”<sup>185</sup>

Even if judges believe that exclusion is the correct outcome, they may also be influenced by the threat of political backlash. This concern is likely particularly strong for elected judges. Judicial elections both motivate judicial candidates to espouse “tough on crime” views, and may weed out candidates who are not suitably tough on crime.<sup>186</sup> Defense attorneys who run for



judicial seats have been attacked for upholding “the rights of violent criminals,”<sup>187</sup> and incumbent judges have been criticized by opponents for having used “loophole[s]” to \*1566 reduce sentences post-conviction.<sup>188</sup> For example, in the 2014 race for the Arkansas Supreme Court, an ad supporting Judge Robin Wynne touted his “refus[al] to allow technicalities to overturn convictions.”<sup>189</sup> Presumably, *Daubert* reliability is just the kind of “technicality” to which this ad was referring. And judicial elections are more than just a platform for anti-defendant rhetoric: Studies have consistently demonstrated that “the pressures of upcoming re-election and retention election campaigns make judges more punitive toward defendants in criminal cases.”<sup>190</sup> These effects are exacerbated in jurisdictions that have liberal campaign spending laws, and, relatedly, in races with higher levels of TV advertising.<sup>191</sup> Although political effects may be felt particularly strongly by elected judges, these findings from the judicial election context suggest that political pressure overall favors prosecutors and disfavors criminal defendants. In other words, to the extent that a trial judge facing an admissibility determination is influenced by political pressure, it is likely to push her toward admitting the evidence.

## 2. *The Path to Solving the Forensic Science Problem*

The structural limitations that trial judges face in the context of forensic science experts proffered by prosecutors, as well as the legal and policy framework in which they operate, help to explain why judges have not been more aggressive in policing untested or simply invalid forensic science evidence in their courtrooms, despite their affirmative gatekeeping obligation. But while judges may feel hemmed in by the limited information at hand (at least insofar as they are relying on the parties to provide them with information), trial judges, at least in the federal system, have the tools to effectively police shoddy forensic evidence.

Federal Rule of Evidence 702 instructs that a qualified witness may testify as an expert only if “the testimony is based on sufficient facts or data,” “the testimony is the product of reliable principles and \*1567 methods,” and “the expert has reliably applied the principles and methods to the facts of the case.”<sup>192</sup> Pursuant to Federal Rule of Evidence 104, the proponent of the evidence bears the burden of showing by a preponderance of the evidence that it meets the requirements of Rule 702.<sup>193</sup> Absent a sufficient showing, the evidence should be appropriately cabined,<sup>194</sup> or, if no relevant evidence is adequately supported, excluded altogether. Reliability of evidence must be established in each case, and is not a matter of precedent in the federal court system.<sup>195</sup>

It may be difficult for judges to stick their necks out and exclude evidence because of an absence of information. Although at least one forensic discipline has been “debunked” by studies affirmatively demonstrating that it is unreliable—namely bite-mark analysis<sup>196</sup>--and others have been called into such serious question that the FBI has abandoned them,<sup>197</sup> the majority of the criticisms leveled at forensic disciplines is that they’re untested: We simply don’t know whether they are reliable or not.<sup>198</sup> Under Rule 702, the absence of evidence that a forensic methodology is reliable is a legally sufficient basis for exclusion, but it is nonetheless rhetorically weaker than affirmative \*1568 evidence showing, for example, that a methodology has a particularly high rate of error.<sup>199</sup>

Nevertheless, trial judges are obligated to ensure that the expert testimony presented in their courtrooms meets the basic standards articulated in Rule 702, and they have the potential to dramatically affect the landscape of forensic science when they exercise that obligation. For example, a forensic discipline known as comparative bullet lead analysis (CBLA) was once commonly admitted in courts.<sup>200</sup> CBLA was primarily used in cases where bullets were too mutilated to be subjected to the more common “tool marks” or “ballistics” analysis.<sup>201</sup> The technique involves analyzing the elements in bullet lead, and applying defined standards to assess whether the lead from the crime scene and the lead from bullets seized from the defendant are “analytically indistinguishable.”<sup>202</sup> However, analysts had no basis for the conclusion that, if the bullets were indeed “analytically indistinguishable,” that such a finding suggested they came from the same box of bullets, or even the same geographic region.<sup>203</sup>

It was in this context that, in 2003, Judge Ronald Guzmán of the Northern District of Illinois rigorously applied Rule 702 to CBLA, and concluded that the technique came up short.<sup>204</sup> In *United States v. Mikos*, Judge Guzmán was confronted with an absence of data to support the claims of the government’s CBLA expert,<sup>205</sup> and concluded \*1569 on that basis that “the

required standard of scientific reliability” was met only with regard to the statement that the bullets were “analytically indistinguishable,” and not with regard to the “further opinion that from this finding it follows that the bullets must or even likely came from the same batch or melt.”<sup>206</sup> The government’s proffered expert, Charles Peters, contended that one of the elements found in the bullets in question, bismuth, was found in “remarkably unusual” quantities in the two bullets.<sup>207</sup> Judge Guzmán was not persuaded by this characterization:

There is ... absolutely no way to know if the bismuth level in these bullets is in fact “remarkably unusual.” It may be so in Peters’ experience, but as we have pointed out, given the huge population with which we are concerned (in the billions), Mr. Peters’ experience is no more than anecdotal evidence. Such evidence can be particularly misleading because it appears logical and reasonable. If agent Peters has not, in years of experience and after hundreds of analyses, previously encountered such a high quantity of bismuth, then why should we not conclude that this is a highly unusual occurrence? The answer to that question lies in the huge size of the bullet population and the relative insignificance of agent Peters’ own personal experience in such a huge population.<sup>208</sup>

In reaching this conclusion, Judge Guzmán recognized the long history of both FBI reliance on CBLA, and of federal courts admitting CBLA testimony at trial.<sup>209</sup> In spite of this recognition, he followed the mandate of Rule 702 and excluded the unsupported testimony.<sup>210</sup>

Within a year, the National Research Council had issued a report concluding that the CBLA discipline could not support statements about how unusual it was for bullets to be “analytically indistinguishable.”<sup>211</sup> A year after that, the FBI discontinued its CBLA Laboratory.<sup>212</sup> After more than thirty years of judges admitting CBLA evidence, its façade of scientific reliability crumbled rapidly.

\*1570 This analysis is not meant to suggest that a single, unpublished opinion from a district judge will lead to the imminent downfall of a forensic discipline. In fact, in the CBLA context, the National Research Council had begun meeting to prepare their report nearly a year before Judge Guzmán issued the *Mikos* ruling, and did so because CBLA had already “come under greater scrutiny.”<sup>213</sup> However, as the ruling to exclude CBLA evidence, *Mikos* was an important milestone in the path toward recognizing and addressing the forensic science problem in one context,<sup>214</sup> and demonstrates that judges can--and should--leverage Rule 702 to block bad evidence even in the face of a long history of admission.

## CONCLUSION

Trial judges are uniquely well positioned to staunch the flow of unreliable forensic evidence into court, which will both prevent wrongful convictions and inspire scientific research to validate or improve forensic disciplines. In order to do so, however, trial judges must break with sometimes-lengthy histories of admission, engage in a technical analysis outside the wheelhouse of most lawyers, and perhaps even face political backlash against an unpopular decision. As difficult as this may seem, none of the obstacles facing trial judges are insurmountable, and none exempt trial judges from their obligation to vigilantly gatekeep expert evidence in their courtrooms.

Most importantly, if the impetus for change does not come from trial judges, the current political climate suggests that bad forensic evidence will continue to be admitted, and history tells us that wrongful convictions will follow. This Note has suggested a path forward for judges prepared to recognize and act on the forensic science problem. Now more than ever, trial judges must lead the way toward a better future for forensic evidence.

## Footnotes

<sup>a1</sup> Copyright © 2017 by Stephanie L. Damon-Moore. J.D., 2017, New York University School of Law; B.A., 2011, Vassar College.

Thank you to Judge Alex Kozinski for suggestions early in the process, Judge Harry T. Edwards for helpful comments and criticism, and Professor Erin Murphy for advice and support. I am also grateful for the early enthusiasm and ongoing diligence of the *N.Y.U. Law Review* editors, particularly Nick Krafft, Eliana Theodorou, Matthew Tieman, and Jiyae Hwang.

- <sup>1</sup> *E.g.*, PRESIDENT’S COUNCIL OF ADVISORS ON SCI. & TECH., FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016) [hereinafter PCAST REPORT], [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensic\\_science\\_report\\_final.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf) (report authored by “an advisory group of the Nation’s leading scientists and engineers”).
- <sup>2</sup> *E.g.*, *Gimenez v. Ochoa*, 821 F.3d 1136, 1144 & n.4 (9th Cir. 2016) (Kozinski, J.) (recognizing that we are in an “age where forensics that were once considered unassailable are subject to serious doubt” and collecting sources); Harry T. Edwards & Jennifer L. Mnookin, Opinion, *A Wake-Up Call on the Junk Science Infesting Our Courtrooms*, WASH. POST (Sept. 20, 2016), [https://www.washingtonpost.com/opinions/a-wake-up-call-on-the-junk-science-infesting-our-courtrooms/2016/09/19/85b6eb22-7e90-11e6-8d13-d7c704ef9fd9\\_story.html](https://www.washingtonpost.com/opinions/a-wake-up-call-on-the-junk-science-infesting-our-courtrooms/2016/09/19/85b6eb22-7e90-11e6-8d13-d7c704ef9fd9_story.html) (opinion piece by a D.C. Circuit Judge and the Dean of UCLA School of Law); Nancy Gertner, Opinion, *Judges Need to Set a Higher Standard for Forensic Evidence*, N.Y. TIMES: ROOM FOR DEBATE (Feb. 4, 2016, 6:36 PM), <http://www.nytimes.com/roomfordebate/2015/03/30/robert-durst-handwriting-and-judging-forensic-science/judges-need-to-set-a-higher-standard-for-forensic-evidence> (editorial authored by a former federal district court judge).
- <sup>3</sup> *See, e.g.*, Keith A. Findley, *Reforming the ‘Science’ in Forensic Science*, 88 WISC. LAW. 32 (2015); Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1 (2009); Jennifer L. Mnookin, *The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate*, 7 LAW, PROBABILITY & RISK 127 (2008).
- <sup>4</sup> *See, e.g.*, Press Release, FBI, FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review (Apr. 20, 2015) [hereinafter Microscopic Hair Analysis Press Release], <https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review> (quoting Peter Neufeld, Co-Director of the Innocence Project, concluding that “FBI microscopic hair analysts committed widespread, systematic error, grossly exaggerating the significance of their data under oath with the consequence of unfairly bolstering the prosecutions’ case”).
- <sup>5</sup> *See* NAT’L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009) [hereinafter NRC REPORT], <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>; PCAST REPORT, *supra* note 1; Craig M. Cooley, *Nurturing Forensic Science: How Appropriate Funding and Government Oversight Can Further Strengthen the Forensic Science Community*, 17 TEX. WESLEYAN L. REV. 441 (2011); Garrett & Neufeld, *supra* note 3.
- <sup>6</sup> *See* Mary A. Bush et al., *Statistical Evidence for the Similarity of the Human Dentition*, 56 J. FORENSIC SCI. 118 (2011) (debunking a foundational assumption of bitemark analysis that human mouths are unique); *see also* PCAST REPORT, *supra* note 1, at 83-87 (surveying the limited research on bite-mark analysis and concluding that “the observed false positive rates were so high that the method is clearly scientifically unreliable at present”).
- <sup>7</sup> *See* Matthew Shaer, *The False Promise of DNA Testing*, ATLANTIC, June 2016, <https://www.theatlantic.com/magazine/archive/2016/06/a-reasonable-doubt/480747/> (describing how DNA testing has become less reliable as scientists rely on increasingly smaller samples and larger numbers of contributors); *see also* ERIN E. MURPHY, INSIDE THE CELL: THE DARK SIDE OF FORENSIC DNA 29-150 (2015) (describing problems with both DNA analysis and testimony by DNA analysts).
- <sup>8</sup> *Misapplication of Forensic Science, INNOCENCE PROJECT*, <https://www.innocenceproject.org/causes/misapplication-forensic-science/> (last visited July 7, 2017).

- <sup>9</sup> See, e.g., NRC REPORT, *supra* note 5, at 106 (“Review of reported judicial opinions reveals that, at least in criminal cases, forensic science evidence is not routinely scrutinized pursuant to the [applicable] standard of reliability ....”); Erin Murphy, *Neuroscience and the Civil/Criminal Daubert Divide*, 85 *FORDHAM L. REV.* 619, 624 (2016) (“[N]early all of the common forensic techniques ... routinely admitted by courts[ ] have been repeatedly denounced as lacking in any scientific basis.”).
- <sup>10</sup> See, e.g., NRC REPORT, *supra* note 5, at 42 (“Most [forensic] techniques were developed in crime laboratories to aid in the investigation of evidence from a particular crime scene, and researching their limitations and foundations was never a top priority.”).
- <sup>11</sup> See, e.g., David E. Bernstein, *Expert Witnesses, Adversarial Bias, and the (Partial) Failure of the Daubert Revolution*, 93 *IOWA L. REV.* 451, 460 (2008) (noting that forensic labs’ frequent dependence on police departments for their budgets may lead to a “desire to please the police”); Glen Whitman & Roger Koppl, *Rational Bias in Forensic Science*, 9 *LAW, PROBABILITY & RISK* 69, 71 (2010) (finding that “tension exists within an institutional structure that frequently puts crime labs under the administration of law enforcement agencies”).
- <sup>12</sup> See, e.g., NRC REPORT, *supra* note 5, at 224-25 (“Commentators have noted repeatedly the deficiencies of forensic science education programs.”); KELLY M. PYREK, *FORENSIC SCIENCE UNDER SIEGE: THE CHALLENGES OF FORENSIC LABORATORIES AND THE MEDICO-LEGAL DEATH INVESTIGATION SYSTEM* 93-96 (2007) (“[A] chronic lack of funding keeps many forensic practitioners from the pursuit of much-needed training.”); Cooley, *supra* note 5, at 450-60 (noting that many forensic analysts are “inadequately trained” in the relevant sciences); Joseph L. Peterson & Anna S. Leggett, *The Evolution of Forensic Science: Progress amid the Pitfalls*, 36 *STETSON L. REV.* 621, 626 (2007) (noting that studies have demonstrated a “severe need for higher-education programs to prepare future forensic scientists for positions in government laboratories”).
- <sup>13</sup> See Stanley Z. Fisher, *In Search of the Virtuous Prosecutor: A Conceptual Framework*, 15 *AM. J. CRIM. L.* 197, 208-11 (1988) (examining structural incentives that push prosecutors to emphasize convictions).
- <sup>14</sup> See, e.g., Peter J. Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggestions for Reform*, 95 *AM. J. PUB. HEALTH* S107, S110 (2005); see also *infra* Section II.B (discussing challenges defense attorneys face in challenging bad forensic evidence proffered by prosecutors).
- <sup>15</sup> See *infra* notes 145-46; see also Neufeld, *supra* note 14, at S109 (reviewing criminal and civil decisions and finding that judges excluded evidence proffered by civil plaintiffs and criminal defendants much more often than evidence proffered by prosecutors). But see Jennifer L. Groscup & Steven D. Penrod, *Battle of the Standards for Experts in Criminal Cases: Police vs. Psychologists*, 33 *SETON HALL L. REV.* 1141, 1144-46 (2003) (surveying the literature on judicial understanding of *Daubert* factors and finding that judges may not understand them all).
- <sup>16</sup> See *infra* Part I (describing some of the most basic and most common problems with forensic evidence).
- <sup>17</sup> See, e.g., *infra* notes 31-34 and accompanying text (discussing testimony by bitemark experts).
- <sup>18</sup> See, e.g., Findley, *supra* note 3, at 39 (noting that “federal courts applying *Daubert* almost never exclude prosecution-proffered forensic science evidence,” but do exclude evidence proffered by criminal defendants and civil plaintiffs, and that the reason for this disparate treatment “is not fully understood”); Jane Campbell Moriarty, “*Misconvictions*,” *Science, and the Ministers of Justice*, 86 *NEB. L. REV.* 1, 37 (2007) (noting that “trial courts have steadfastly refused to take the *Daubert* trilogy language seriously as applied to [certain forensic] evidence,” without addressing why that is the case); Jane Campbell Moriarty & Michael J.



Saks, *Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping*, JUDGES' J., Fall 2005, at 16, 28-29 (noting that judges fail to uphold gatekeeping responsibilities in both *Frye* and *Daubert* jurisdictions, without taking up the source of this failure); Michael J. Saks & Ashley M. Votruba, "... *And the Courts Have Been Utterly Ineffective*," JUDGES' J., Summer 2015, at 28, 29-30 (same). *But see* Julie A. Seaman, *A Tale of Two Dauberts*, 47 GA. L. REV. 889, 900, 902 (2013) (suggesting that variation in the quality of experts and, reluctance in criminal cases, to "exclude prosecution evidence that carries a long historical pedigree" could explain admission of handwriting analysis evidence in criminal cases at higher rates than civil cases).

<sup>19</sup> *Criminology & Criminal Justice Studies*, UNIV. MICH.-DEARBORN, <https://umdearborn.edu/casl/undergraduate-programs/areas-study/criminology-criminal-justice-studies> (last visited July 9, 2017).

<sup>20</sup> Donald E. Shelton, *Forensic Science Evidence and Miscarriages of Justice*, in ADVANCES IN FORENSIC HUMAN IDENTIFICATION 409, 417-21 (Xanthé Mallett et al. eds., 2014) [hereinafter Shelton, *Miscarriages of Justice*]; Donald E. Shelton, *Forensic Science Evidence and Judicial Bias in Criminal Cases*, JUDGES' J., Summer 2010, at 18, 22-24 [hereinafter Shelton, *Judicial Bias*]; *see also* Adam B. Shneiderman, *Prosecutors Respond to Calls for Forensic Science Reform: More Sharks in Dirty Water*, 126 YALE L.J.F. 348, 353-56 (2017) (attributing "the judiciary's failure to keep bad science out of courtrooms" to "a systemic pro-prosecution bias on the bench").

<sup>21</sup> *See infra* notes 157-71 (discussing the influence of biases and heuristics on admissibility decisions).

<sup>22</sup> *See infra* Section III.A.1 (discussing the problem that judges commonly lack scientific training and may struggle to rigorously assess reliability of expert evidence as a result).

<sup>23</sup> *See infra* notes 149-56 and accompanying text (discussing the relative absence of effective defense challenges to prosecution experts).

<sup>24</sup> *See infra* notes 173-85 and accompanying text (describing how the Federal Rules of Evidence as well as broader policy concerns may lead judges to err on the side of admitting untested evidence).

<sup>25</sup> As Blackstone said, and many a legal commentator has echoed, it is "better that ten guilty persons escape, than that one innocent suffer." 4 WILLIAM BLACKSTONE, COMMENTARIES \*352. Whether we can reasonably infer anything from the widespread adoration of this sentiment in the legal community is an entirely different question. *See generally* Alexander Volokh, n *Guilty Men*, 146 U. PA. L. REV. 173 (1997) (providing a cheeky analysis of the various versions of this sentiment).

<sup>26</sup> The only forensic science discipline that has generally escaped criticism is single-source and simple-mixture DNA testing. Compare PCAST REPORT, *supra* note 1 (finding serious flaws across nearly all disciplines evaluated, including DNA analysis of complex-mixture samples, bite-mark analysis, latent fingerprint analysis, firearms analysis, footwear analysis, and hair analysis), with *id.* at 69-73 (describing single-source and simple-mixture DNA analysis as "objective methods whose foundational validity has been properly established," but cautioning that these methodologies are "not infallible in practice").

<sup>27</sup> For detailed analyses of the shortcomings of forensic evidence, see generally PCAST REPORT, *supra* note 1, and NRC REPORT, *supra* note 5. *See also* Garrett & Neufeld, *supra* note 3 (cataloguing problems with forensic testimony specifically).

<sup>28</sup> PCAST actually advises against permitting experts to testify to a "match," as the term is "likely to imply an inappropriately high probative value." PCAST REPORT, *supra* note 1, at 45-46. They recommend the use of the phrase "proposed identification" instead. *Id.* at 46.

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

---

- <sup>29</sup> See AM. BD. OF FORENSIC ODONTOLOGY, DIPLOMATES REFERENCE MANUAL 98-99 (Mar. 2017 ed.), <http://abfo.org/wp-content/uploads/2012/08/ABFO-Reference-Manual-April-2017-v7.pdf>.
- <sup>30</sup> See *id.* at 102.
- <sup>31</sup> See 4 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 35:4 (2016-2017 ed. 2016) (noting the irony that although “forensic odontologists, perhaps reflecting a grounding in scientific skepticism that is often absent from the more traditional forensic identification sciences, were themselves somewhat doubtful about whether the state of their knowledge permitted them to successfully identify a perpetrator ‘to the exclusion of all others,’” admission by courts apparently convinced them “that, despite their doubts, they were indeed able to perform bitemark identifications after all” (internal citation omitted)).
- <sup>32</sup> 600 S.W.2d 541, 563 (Mo. Ct. App. 1980) (admitting bite-mark analysis testimony despite lack of standardized methodology).
- <sup>33</sup> *Stubbs v. State*, 845 So. 2d 656, 669 (Miss. 2003) (quoting *Howard v. State*, 701 So. 2d 274, 288 (Miss. 1997)).
- <sup>34</sup> *Sager*, 600 S.W.2d at 563 (emphasis added). Though the claim that human mouths are unique goes to the heart of forensic odontology, that claim was debunked in a 2010 study. Bush et al., *supra* note 6.
- <sup>35</sup> See PCAST REPORT, *supra* note 1, at 84-85.
- <sup>36</sup> In addition to the pernicious problems in forensic odontology, commentators have also had a field day with bite-related puns. *E.g.*, Erica Beecher-Monas, *Reality Bites: The Illusion of Science in Bite-Mark Evidence*, 30 CARDOZO L. REV. 1369 (2009); Adam Deitch, Comment, *An Inconvenient Tooth: Forensic Odontology Is an Inadmissible Junk Science When It Is Used to “Match” Teeth to Bitemarks in Skin*, 2009 WISC. L. REV. 1205.
- <sup>37</sup> See Lyn Haber & Ralph Norman Haber, *Scientific Validation of Fingerprint Evidence Under Daubert*, 7 LAW, PROBABILITY & RISK 87, 105 (2008) (finding no support for the scientific validity of the most common method of fingerprint analysis); Mnookin, *supra* note 3, at 131 (noting the “near-universal” judicial acceptance of fingerprint analysis, and her own conclusion that most fingerprint evidence should be excluded under *Daubert*). Of course, today the “gold standard” forensic discipline is DNA evidence, which has recently garnered some criticism of its own. See *supra* note 7; see also PCAST REPORT, *supra* note 1, at 78-86 (describing problems with DNA analysis of complex-mixture samples).
- <sup>38</sup> FBI, THE SCIENCE OF FINGERPRINTS: CLASSIFICATIONS AND USES iv (1984).
- <sup>39</sup> See Mnookin, *supra* note 3, at 128 (describing judicial opinions that laud fingerprint evidence as having “survived an entire century of testing within the crucible of the courtroom”).
- <sup>40</sup> NRC REPORT, *supra* note 5, at 140-41 (discussing the degree to which latent fingerprint analysis relies on the subjective interpretation of individual examiners).
- <sup>41</sup> PCAST REPORT, *supra* note 1, at 91.
- <sup>42</sup> *Id.* at 88. The PCAST Report also concludes that “estimated false positive rates are *much higher* than the general public ... would

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

---

likely believe based on longstanding claims about the accuracy of fingerprint analysis.” *Id.* at 95.

<sup>43</sup> Garrett & Neufeld, *supra* note 3, at 36 (citing Dale D. Dykes, *The Use of Frequency Tables in Parentage Testing*, in PROBABILITY OF INCLUSION IN PATERNITY TESTING: A TECHNICAL WORKSHOP 15, 20, 29 (Herbert Silver ed., 1982)).

<sup>44</sup> *Id.* at 49.

<sup>45</sup> *Id.* at 53-54.

<sup>46</sup> Microscopic Hair Analysis Press Release, *supra* note 4.

<sup>47</sup> Such a decision would not only solve the problem that is the focus of this Note--wrongful convictions based on junk science--but, by eliminating demand for “junk science” and creating demand for reliable evidence, would also provide a massive incentive for the forensic science community to adopt scientific approaches to forensic testing, and abandon disciplines that could not meet high standards.

<sup>48</sup> MODEL RULES OF PROF’L CONDUCT r. 3.8 cmt. 1 (AM. BAR ASS’N 2014) (declaring that “[a] prosecutor has the responsibility of a minister of justice and not simply that of an advocate,” which requires prosecutors to “see that the defendant is accorded procedural justice,” ensure “that guilt is decided upon the basis of sufficient evidence,” and take “special precautions ... to prevent and rectify the conviction of innocent persons”).

<sup>49</sup> See Fisher, *supra* note 13, at 208-11 (describing ways in which the adversarial system leads prosecutors to “invest [their] energies single-mindedly in maximizing convictions and punishments”); see also Stephanos Bibas, *Prosecutorial Regulation Versus Prosecutorial Accountability*, 157 U. PA. L. REV. 959, 985 & n.98 (2009) (describing how media coverage favors prosecutors with high conviction rates).

<sup>50</sup> See, e.g., Kenneth Bresler, Essay, “*I Never Lost a Trial*”: When Prosecutors Keep Score of Criminal Convictions, 9 GEO. J. LEGAL ETHICS 537, 541-44 (1996) (discussing how the practice of some prosecutors to tally their convictions violates “[t]he most notable and noble principle of prosecution ... to seek justice, not convictions”); Fred C. Zacharias, *Structuring the Ethics of Prosecutorial Trial Practice: Can Prosecutors Do Justice?*, 44 VAND. L. REV. 45, 107-08 (1991) (arguing that the “do justice” mandate is hamstrung by the adversarial nature of criminal prosecutions where “[w]inning is at a premium” and “competitive juices flow”).

<sup>51</sup> See, e.g., DEBORAH TUERKHEIMER, *FLAWED CONVICTIONS: ‘SHAKEN BABY SYNDROME’ AND THE INERTIA OF JUSTICE* 148 (2014) (describing a shaken baby syndrome (SBS) case where the prosecutor dismissed the charges after the state’s own experts disavowed a shaken baby syndrome diagnosis, but stated the office “wholeheartedly” believed the defendant was guilty nonetheless); Shaila Dewan, *Despite DNA Test, a Case Is Retried*, N.Y. TIMES (Sept. 6, 2007), <http://www.nytimes.com/2007/09/06/us/06dna.html> (describing District Attorney Forrest Allgood’s decision to retry a rape and murder case after the convicted man, Kennedy Brewer, was excluded by DNA); see also Radley Balko, Opinion, *Election Results: One of America’s Worst Prosecutors Lost Last Night, but One of Its Worst Attorney Generals Won*, WASH. POST (Nov. 4, 2015), <https://www.washingtonpost.com/news/the-watch/wp/2015/11/04/election-results-one-of-americas-worst-prosecutors-lost-last-night-but-one-of-its-worst-attorneys-general-won/> (noting that Kennedy Brewer spent an additional seven years in prison after DNA exonerated him because of District Attorney Forrest Allgood’s decision to re-prosecute him for the crime).

<sup>52</sup> SANDRA GUERRA THOMPSON, *COPS IN LAB COATS: CURBING WRONGFUL CONVICTIONS THROUGH INDEPENDENT FORENSIC LABORATORIES* 109 (2015) (describing the development of forensic science as a tool for

investigating and prosecuting crimes). As an investigatory tool, untested or unreliable methods may still be helpful. *See* PCAST REPORT, *supra* note 1, at 4 (noting that forensic science used for investigative purposes, as opposed to prosecution purposes, may appropriately “come from both well-established science and exploratory approaches”). But untested methods that are worth a shot when investigating a case can return wrongful convictions if introduced in court.

53 *See* NAT’L CONFERENCE OF STATE LEGISLATURES, ADMINISTRATION OF FORENSIC SCIENCE DEPARTMENTS 1 (2013), <http://www.ncsl.org/Documents/cj/AdministrationOfForensicServices.pdf> (noting that only six states--Alabama, Arkansas, Connecticut, Delaware, Rhode Island, and Virginia--and the District of Columbia have forensic laboratories “that operate as ... independent agenc[ies] or as ... department[s] ... that do[ ] not share oversight with prosecutorial or law enforcement services”). For a comprehensive analysis of the ways in which law enforcement interests have obstructed efforts to fund and perform independent scientific testing of forensic science, see Paul C. Giannelli, **Daubert** and *Forensic Science: The Pitfalls of Law Enforcement Control of Scientific Research*, 2011 U. ILL. L. REV. 53.

54 *See* Brandon L. Garrett, *The Genetic Panopticon*, BOS. REV. (Mar. 17, 2016), <http://bostonreview.net/us/brandon-garrett-dna-crime-lab-forensics> (describing police and prosecutors as the “chief consumers of forensic evidence”).

55 As Professor Deborah Tuerkheimer points out, the doctor whose paper served as the foundation for SBS, A. Norman Guthkelch, has been a vocal critic of SBS prosecutions, writing that there was “not a vestige of proof when the name [SBS] developed that shaking alone causes the triad [of symptoms said to characterize SBS].” TUERKHEIMER, *supra* note 51, at 87.

56 *See id.* at 2 (noting that only fifteen SBS cases reached the appeals courts before 1990, more than two hundred were appealed between 1990 and 2000, and more than eight hundred were appealed between 2000 and 2010).

57 *Id.* at 35-37 (discussing the development of child abuse units and the impact these units have on SBS prosecutions).

58 *Id.* at xi.

59 *Id.*

60 *Id.* at 17.

61 *See* Lauryn Schroeder, *Pinpointing Shaken-Baby Syndrome Cases: A New Medill Justice Project Study Identifies Where Higher Rates of Shaken-Baby Syndrome Cases Are Occurring in the United States*, MEDILL JUST. PROJECT (Dec. 10, 2013), <http://www.medilljusticeproject.org/2013/12/10/hot-spots/> (suggesting that these prosecution “hot spots” are, in part, the result of a collaboration among medical professionals and prosecutors).

62 *Id.*

63 *Id.*

64 PCAST REPORT, *supra* note 1.

65 Letter from Michael A. Ramos, President, Nat’l Dist. Attorneys Ass’n, to President Barack Obama 8 (Nov. 16, 2016),



<http://www.ciclt.net/ul/ndaajustice/PCAST/NDAA%20PCAST%20Response%20FINAL.pdf>. These are choice words from the NDAA given the PCAST Report's emphatic argument that experience cannot be used as a basis to determine the validity of forensic disciplines. PCAST REPORT, *supra* note 1, at 32-33 ("Casework is not scientifically valid research, and experience alone cannot establish scientific validity.").

<sup>66</sup> Letter from Michael A. Ramos to President Barack Obama, *supra* note 65, at 8.

<sup>67</sup> Gary Fields, *White House Advisory Council Report Is Critical of Forensics Used in Criminal Trials: U.S. Attorney General Says Justice Department Won't Adopt Recommendations*, WALL ST. J. (Sept. 20, 2016, 4:25 PM), <https://www.wsj.com/articles/white-house-advisory-council-releases-report-critical-of-forensics-used-in-criminal-trials-1474394743> (quoting statement of U.S. Attorney General Loretta Lynch).

<sup>68</sup> Spencer S. Hsu, *Sessions Orders Justice Dept. to End Forensic Science Commission, Suspend Review Policy*, WASH. POST (Apr. 10, 2017), [https://www.washingtonpost.com/local/public-safety/sessions-orders-justice-dept-to-end-forensic-science-commission-suspend-review-policy/2017/04/10/2dada0ca-1c96-11e7-9887-1a5314b56a08\\_story.html](https://www.washingtonpost.com/local/public-safety/sessions-orders-justice-dept-to-end-forensic-science-commission-suspend-review-policy/2017/04/10/2dada0ca-1c96-11e7-9887-1a5314b56a08_story.html).

<sup>69</sup> *See id.*

<sup>70</sup> *Id.*

<sup>71</sup> Letter from Comm'rs, NCFS, to Hon. Jefferson B. Sessions, U.S. Attorney Gen. & Dr. Kent Rochford, Acting Dir. of Nat'l Inst. of Standards and Tech. (Apr. 6, 2017), <https://assets.documentcloud.org/documents/3549346/Scientists-on-national-commission-urge-panel-be.pdf>.

<sup>72</sup> *Id.*

<sup>73</sup> Sally Q. Yates, Deputy Attorney Gen., Remarks at the Ninth Meeting of the National Commission on Forensic Science (Mar. 21, 2016), <https://www.justice.gov/opa/speech/deputy-attorney-general-sally-q-yates-delivers-remarks-ninth-meeting-national-commission>.

<sup>74</sup> Hsu, *supra* note 68.

<sup>75</sup> *Id.*

<sup>76</sup> Radley Balko, Opinion, *When Obama Wouldn't Fight for Science*, WASH. POST (Jan. 4, 2017), [https://www.washingtonpost.com/news/the-watch/wp/2017/01/04/when-obama-wouldnt-fight-for-science/?utm\\_term=.2e6515f74e0c](https://www.washingtonpost.com/news/the-watch/wp/2017/01/04/when-obama-wouldnt-fight-for-science/?utm_term=.2e6515f74e0c).

<sup>77</sup> Letter from Michael A. Ramos to President Barack Obama, *supra* note 65, at 8.

<sup>78</sup> *See, e.g.*, Neufeld, *supra* note 14, at S110 (noting that although there were nearly fifteen million criminal charges filed in state courts in the year 2000, a review of published opinions from state criminal courts from August 1999 through August 2000 turned up only fifty defense challenges to admissibility).

- <sup>79</sup> Alexa Van Brunt, *Poor People Rely on Public Defenders Who Are Too Overworked to Defend Them*, GUARDIAN (June 17, 2015, 7:30 AM), <https://www.theguardian.com/commentisfree/2015/jun/17/poor-rely-public-defenders-too-overworked> (noting that more than eighty percent of people facing felony charges are indigent and rely on representation by a public defender); *see also* CAROLINE WOLF HARLOW, BUREAU OF JUSTICE STATISTICS, DEFENSE COUNSEL IN CRIMINAL CASES 1 (2000), <https://www.bjs.gov/content/pub/pdf/dccc.pdf> (“At felony case termination, court-appointed counsel represented 82% of State defendants in the 75 largest counties in 1996 and 66% of Federal defendants in 1998.”).
- <sup>80</sup> *See* Jaeah Lee, Hannah Levintova & Brett Brownell, *Charts: Why You’re in Deep Trouble if You Can’t Afford a Lawyer*, MOTHER JONES (May 6, 2013, 10:00 AM), <http://www.motherjones.com/politics/2013/05/public-defenders-gideon-supreme-court-charts> (comparing the recommended caseload numbers to the reality across cases of different levels of complexity and concluding that public defenders are overburdened).
- <sup>81</sup> *See* Neufeld, *supra* note 14, at S110 (“[E]ven if inclined to mount a **Daubert** challenge, [criminal defense attorneys] lack the requisite knowledge and skills, as well as the funds, to succeed.”).
- <sup>82</sup> *See* Peter A. Joy & Kevin C. McMunigal, *Does the Lawyer Make a Difference? Public Defender v. Appointed Counsel*, 27 CRIM. JUST. 46, 47 (2012) (describing results of a Rand study which found that court-appointed lawyers in Philadelphia earn about two dollars per hour, which “fail[s] to attract qualified lawyers, discourage[s] adequate preparation, and create[s] an incentive for appointed lawyers to take on many more cases than they can adequately handle”).
- <sup>83</sup> *See, e.g., Does Equal Justice for All Include the Poor?*, NPR (Nov. 5, 2013, 11:49 AM), <http://www.npr.org/templates/story/story.php?storyId=243213638> (quoting Professor Eve Primus discussing various systems of indigent defense, including appointment of counsel who “may or may not have any experience in criminal law”). Despite the overwhelming caseloads that burden public defenders, studies have shown that outcomes are better for defendants with public defenders than those who receive appointed counsel. *See, e.g.,* Joy & McMunigal, *supra* note 82, at 46 (describing a Rand study of the Philadelphia area, which found that public defender representation compared to appointed counsel representation reduced a murder defendant’s conviction rate by nineteen percent, reduced the likelihood of a life sentence by sixty-two percent, and decreased the expected prison term by twenty-four percent).
- <sup>84</sup> *See* Neufeld, *supra* note 14, at S110 (“Even when the most vulnerable forensic sciences--hair microscopy, bite marks, and handwriting--are attacked, the courts routinely affirm admissibility citing earlier decisions rather than facts established at a hearing.”).
- <sup>85</sup> *See infra* notes 142-45 and accompanying text (describing studies finding that judges reject defense proffers at substantially greater rates than proffers from prosecutors or civil parties).
- <sup>86</sup> *See Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 596 (1993) (“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.”).
- <sup>87</sup> *See* Caroline T. Parrott et al., *Differences in Expert Witness Knowledge: Do Mock Jurors Notice and Does It Matter?*, 43 J. AM. ACAD. PSYCHIATRY & L. 69, 77 (2015) (hypothesizing that jurors may be influenced by “the peripheral cue” that “being an expert extend[s] a blanket influence of knowledge,” and may assume that the judge would “allow[ ] only qualified people with specialized knowledge to take the role of expert”); *see also* FED. R. EVID. 702 advisory committee’s note to 2000 amendments (advising courts not to use the word “expert” in order to avoid “put[ting] their stamp of authority” on the testimony) (quoting Charles R. Richey, *Proposals to Eliminate the Prejudicial Effect of the Use of the Word “Expert” Under the Federal Rules of*

*Evidence in Civil and Criminal Jury Trials*, 154 F.R.D. 537, 559 (1994)).

- 88 See *United States v. Frazier*, 387 F.3d 1244, 1263 (11th Cir. 2004) (“Simply put, expert testimony may be assigned talismanic significance in the eyes of lay jurors ....”).
- 89 For an interesting account of an appointed lawyer with no criminal experience, see *Deep End of the Pool*, THIS AMERICAN LIFE (Aug. 28, 2016), <https://www.thisamericanlife.org/radio-archives/episode/595/transcript>.
- 90 Garrett & Neufeld, *supra* note 3, at 89.
- 91 Bert Black et al., *Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge*, 72 TEX. L. REV. 715, 789 (1994) (noting that “most commentators believe ostensibly scientific testimony may sway a jury even when as science it is palpably wrong”).
- 92 See ABA STANDING COMM. ON LEGAL AID & INDIGENT DEFENDANTS, *Gideon’s Broken Promise: America’s Continuing Quest for Equal Justice* 38 (2004) (finding that funding for indigent defense counsel is “shamefully inadequate,” leaving attorneys without “bare necessities for an adequate defense” including experts); Garrett & Neufeld, *supra* note 3, at 89-90 (“[C]ourts frequently deny the defense funding for experts in criminal cases in which forensic evidence plays a central role.”); Neufeld, *supra* note 14, at S110 (“Unlike prosecutors with free access to government medical examiners and publicly funded crime labs, defense counsel must usually seek independent contractors, and then, if the client is indigent, only with the court’s permission.”).
- 93 See Paul C. Giannelli, *Ake v. Oklahoma: The Right to Expert Assistance in a Post-Daubert, Post-DNA World*, 89 CORNELL L. REV. 1305, 1327, 1331 (2004) (noting that while prosecutors have access to “state, county, regional, or metropolitan crime laboratories,” more than half of these laboratories will not test evidence for defense attorneys, giving prosecutors an “overwhelming advantage” in securing expert assistance).
- 94 Groscup & Penrod, *supra* note 15, at 1155, 1165 tbl.1 (in a study of more than 1800 cases, finding that judges admitted 88.6% of prosecution experts and 24.4% of defense experts).
- 95 Lora M. Levett & Margaret Bull Kovera, *Psychological Mediators of the Effects of Opposing Expert Testimony on Juror Decisions*, 15 PSYCHOL. PUB. POL’Y & L. 124, 127 (2009). The “hired gun” heuristic may be applied more often to defense experts than to those proffered by the state, since defense counsel are more likely to rely on “experts for hire” rather than state employees whose compensation is not directly linked to trial testimony. Studies have suggested that, particularly where the testimony in question is complex, juries may rely on peripheral cues, including how often experts testify and how much they are compensated, to weigh the value of their testimony, rather than the content of the testimony itself. See, e.g., Joel Cooper & Isaac M. Neuhaus, *The “Hired Gun” Effect: Assessing the Effect of Pay, Frequency of Testifying, and Credentials on the Perception of Expert Testimony*, 24 LAW & HUM. BEHAV. 149, 162 (2000) (finding that mock jurors perceived highly compensated experts who testified frequently to be “less likeable, honest, trustworthy, [and] believable”).
- 96 Shelton, *Judicial Bias*, *supra* note 20, at 22-23 (providing the results of an ABA survey of the expert evidence tests in all fifty states).
- 97 FED. R. EVID. 702; see also FED. R. EVID. 702 advisory committee’s note to 2000 amendments (noting that Rule 702 was amended after *Daubert* to “affirm[ ] the trial court’s role as gatekeeper”).

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

---

- 98 In fact, at least one federal judge has granted a habeas petition on an ineffective assistance of counsel theory because a criminal defense attorney failed to make a **Daubert** motion. *See* United States v. Hebshie, 754 F. Supp. 2d 89, 120-22 (D. Mass. 2010).
- 99 Because of this obligation, the forensic science problem cannot exclusively--or even primarily--be attributed to non-judicial actors.
- 100 509 U.S. 579 (1993).
- 101 Shelton, *Judicial Bias*, *supra* note 20, at 22-23 (showing that thirty-five states apply the **Daubert** test in at least some circumstances).
- 102 293 F. 1013 (D.C. Cir. 1923).
- 103 *Id.* at 1014. A sizeable minority of states continue to follow the *Frye* test for some or all of their evidentiary decisions, while a few states, including Utah, Virginia, and Wisconsin, use neither **Daubert** nor *Frye*. Shelton, *Judicial Bias*, *supra* note 20, at 22-23. As the dominant rule, **Daubert** is the focus of this Note. However, it is worth noting that much of forensic science ought to be barred under *Frye*'s "general acceptance" standard. The most obvious example of forensic evidence that falls short of this standard is what has been described as "maverick" forensics, evidence offered by experts who "claim to have techniques that have not been heard of or tested, that are based on nothing but the witness's say-so, that others in the field believe do not exist, and that can be performed only by the maverick." Moriarty & Saks, *supra* note 18, at 29.
- 104 **Daubert**, 509 U.S. at 590.
- 105 *Id.* at 593-94. In *Kumho Tire Co. v. Carmichael*, the Supreme Court made clear that **Daubert's** gatekeeping obligation applied to all forms of Rule 702 evidence, not only scientific evidence. 526 U.S. 137, 147 (1999).
- 106 **Daubert**, 509 U.S. at 594 & n.12.
- 107 *Kumho Tire*, 526 U.S. at 150 (quoting **Daubert**, 509 U.S. at 593).
- 108 *Id.* (quoting Brief for United States as Amicus Curiae Supporting Petitioners at 19, *Kumho Tire*, 526 U.S. 137 (No. 97-1709)).
- 109 *Weisgram v. Marley Co.*, 528 U.S. 440, 455 (2000) (noting that **Daubert** put parties on notice of the standards required for admissibility of evidence).
- 110 The **Daubert** trilogy includes: **Daubert**, 509 U.S. 579, *General Electric Co. v. Joiner*, 522 U.S. 136, 143 (1997) (establishing that admissibility rulings are reviewed for abuse of discretion only), and *Kumho Tire*, 526 U.S. 137 (extending **Daubert's** holding to nonscientific experts).
- 111 *See Kumho Tire*, 526 U.S. at 152 ("The trial court must have ... latitude in deciding *how* to test an expert's reliability ... and to decide whether or when special briefing or other proceedings are needed to investigate reliability ....").
- 112 *See id.* at 150.



TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

---

113 *See Joiner*, 522 U.S. at 142.

114 *Kumho Tire*, 526 U.S. at 159 (Scalia, J., concurring).

115 *See, e.g., Murphy*, *supra* note 9, at 621 (noting that *Daubert* was “heralded as a watershed moment in the treatment of scientific evidence” that many commentators believed could “stem the perceived epidemic of ‘junk science’ that had overtaken the courts”); Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 894-95 (2005) (citing *Daubert* as one reason for their predicted paradigm shift); *see also* Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 57 (1998) (arguing that the *Daubert* decision presented an opportunity for courts to begin screening out junk forensic evidence).

116 *United States v. Llera Plaza (Llera Plaza I)*, 179 F. Supp. 2d 492 (E.D. Pa. 2002), *vacated* 188 F. Supp. 2d 549 (E.D. Pa. 2002); *United States v. Llera Plaza (Llera Plaza II)*, 188 F. Supp. 2d 549 (E.D. Pa. 2002).

117 *Llera Plaza I*, 179 F. Supp. 2d at 503 (quoting *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593 (1993)).

118 *Id.* at 506. Judge Pollack pointed out that “to rely on [the adversary process] would be to vitiate the gatekeeping role of federal trial judges, thereby undermining the essence of Rule 702 as interpreted by the Court in *Daubert*.” *Id.*

119 *Id.* at 509.

120 *Id.*

121 *Id.* at 516.

122 *Id.* at 515.

123 *Id.* This critique was largely confirmed by the NRC Report in 2009. NRC REPORT, *supra* note 5, at 142-45.

124 *Llera Plaza I*, 179 F. Supp. 2d at 516.

125 *Id.*

126 *Llera Plaza II*, 188 F. Supp. 2d 549, 552-53, 576 (E.D. Pa. 2002).

127 *Llera Plaza I*, 179 F. Supp. 2d at 516 (“[T]he court finds that the information of record is unpersuasive, one way or another, as to [the methodology’s] ‘scientific’ rate of error ....”).

128 *Llera Plaza II*, 188 F. Supp. 2d at 566. Judge Pollack thus concluded despite crediting the defense expert’s criticisms of the FBI’s proficiency tests. “On the record before me, the FBI examiners got very high proficiency grades, but the tests they took did not.” *Id.* at 565.

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

---

- 129 FED. R. EVID. 702 advisory committee's note to 2000 amendments (citing *Bourjaily v. United States*, 483 U.S. 171 (1987)). A contrary rule would have the absurd result of requiring that criminal defendants (the vast majority of whom are indigent) fund proficiency testing for any dicey forensic discipline they hope to keep out of court. Yet *Llera Plaza II* seems to suggest just such a rule. See *Llera Plaza II*, 188 F. Supp. 2d at 566, 572.
- 130 *Llera Plaza II*, 188 F. Supp. 2d at 572 (emphasis added).
- 131 **Daubert** v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1317 (9th Cir. 1995).
- 132 *Id.* at 1317 n.5 (emphasis added) (internal citations omitted).
- 133 Alex Kozinski, Opinion, *Rejecting Voodoo Science in the Courtroom*, WALL ST. J. (Sept. 19, 2016, 7:36 PM), <https://www.wsj.com/articles/rejecting-vooodoo-science-in-the-courtroom-1474328199> ("Among the more than 2.2 million inmates in U.S. prisons and jails, countless may have been convicted using unreliable or fabricated forensic science.").
- 134 See Sophia I. Gatowski et al., *Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-Daubert World*, 25 LAW & HUM. BEHAV. 433, 444-48 (2001) (discussing results of a survey of 400 state court judges which indicated that only four percent of respondents accurately described the relevance of **Daubert's** falsifiability and error rate factors).
- 135 See **Daubert** v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593-94 (1993).
- 136 *Id.* at 593.
- 137 See, e.g., *United States v. Green*, 405 F. Supp. 2d 104, 109 n.7 (D. Mass. 2005) (noting that, while the Association of Firearm and Tool Mark Examiners (AFTE) "publishes a journal that is peer-reviewed by other members of the field," that "'field' consists entirely of individuals who work for law enforcement agencies" as opposed to the field of DNA analysis, which includes "neutral academics as well as law enforcement personnel").
- 138 See, e.g., *United States v. Otero*, 849 F. Supp. 2d 425, 433 (D.N.J. 2012) ("AFTE theory is subject to peer review through submission to and publication by the AFTE Journal of validation studies which test the theory.").
- 139 See *supra* notes 116-30 and accompanying text (analyzing the *Llera Plaza* rulings).
- 140 *Llera Plaza I*, 179 F. Supp. 2d 492, 509 (E.D. Pa. 2002).
- 141 *Llera Plaza II*, 188 F. Supp. 2d 549, 563-64 (E.D. Pa. 2002) ("The fact that fingerprint specialists are not 'scientists,' and hence that the forensic journals in which their writings on fingerprint identification appear are not 'scientific' journals in **Daubert's** peer review sense, does not seem to me to militate against the utility of the identification procedures employed by fingerprint specialists ....").
- 142 D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99, 105 (2000) (noting that only one of the remaining six cases resulted in a reversal based on the unreliability of expert testimony); Neufeld, *supra* note 14, at S109 (analyzing the results of the Risinger study).

- <sup>143</sup> Risinger, *supra* note 142, at 106-07. At the trial court level, in the twelve published opinions addressing defense challenges to prosecution evidence, the evidence was fully admitted in eleven out of the twelve cases and admitted with restrictions in the remaining one. *Id.* at 109. Of the forty-two cases dealing with prosecution challenges to defense proffers, on the other hand, the prosecution prevailed in twenty-eight, or precisely two-thirds of the time. *Id.* at 110.
- <sup>144</sup> Jennifer L. Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 PSYCHOL. PUB. POL'Y & L. 339, 346 (2002) (concluding that “[t]he party for whom the key expert testified was significantly related to admission” at trial and on appeal).
- <sup>145</sup> **Murphy**, *supra* note 9, at 621-24 (surveying the empirical studies supporting a civil/criminal divide).
- <sup>146</sup> Jane Campbell Moriarty, *Will History Be Servitude?: The NAS Report on Forensic Science and the Role of the Judiciary*, 2010 UTAH L. REV. 299, 315; *see also* **Murphy**, *supra* note 9, at 621.
- <sup>147</sup> *See, e.g.*, Neufeld, *supra* note 14, at S110 (“Unlike the extremely well-litigated civil challenges, the criminal defendant’s challenge is usually perfunctory. ... Defense lawyers generally fail to build a challenge with appropriate witnesses and new data.”); Risinger, *supra* note 142, at 135 (finding a “systematic failure to seriously litigate [the weakest kinds of common prosecution-proffered expertise] on the part of the criminal defense bar”). Under those circumstances, judges may be more likely to defer to the prosecutor, who is expected to be more objective than defense counsel. *See supra* note 48 and accompanying text (discussing the prosecutor’s ostensible role as a “minister of justice”).
- <sup>148</sup> *See* David S. Clark, *The Organization of Lawyers and Judges*, in 16 INTERNATIONAL ENCYCLOPEDIA OF COMPARATIVE LAW 3-74 (Mauro Cappelletti ed., 2014) (noting that while judges in civil law systems typically bear the burden for independently researching the law, “the common law system of procedure puts a heavy burden on attorneys to discover the facts in a case and even to brief legal issues for the judge”).
- <sup>149</sup> *See supra* notes 78-85 and accompanying text (discussing why defense attorneys seldom file motions challenging prosecution proffers).
- <sup>150</sup> It is likely that judges, like most people, generally believe that forensic science is valid and reliable. *See* Gretchen Gavett, *Judge Harry T. Edwards: How Reliable Is Forensic Evidence in Court?*, PBS (Apr. 17, 2012), <http://www.pbs.org/wgbh/frontline/article/judge-harry-t-edwards-how-reliable-is-forensic-evidence-in-court/> (quoting D.C. Circuit Judge Harry Edwards’s statement that he and his colleagues “assumed that the methodology [of forensic science] was valid and reliable and that the work in putting the evidence together and in offering the testimony was proper” until the NRC uncovered “systemic, serious problems with respect to certain of the [forensic] disciplines”).
- <sup>151</sup> *See supra* notes 92-95 and accompanying text (discussing obstacles defense attorneys experience to obtaining experts, particularly in forensic fields).
- <sup>152</sup> NRC REPORT, *supra* note 5 (describing the various types of forensic science, standards of admissibility as evidence in legal cases, and problems in methodology and oversight).
- <sup>153</sup> PCAST REPORT, *supra* note 1 (evaluating the scientific validity of seven types of forensic evidence, and providing specific recommendations for the judiciary).

- <sup>154</sup> See *supra* note 3 (collecting sources that critique forensic evidence).
- <sup>155</sup> See, e.g., Balko, *supra* note 76 (discussing problems with forensic science and Obama's failure to take action to solve them); Spencer S. Hsu, *FBI Admits Flaws in Hair Analysis over Decades*, WASH. POST (Apr. 18, 2015), [https://www.washingtonpost.com/local/crime/fbi-overstated-forensic-hair-matches-in-nearly-all-criminal-trials-for-decades/2015/04/18/39c8d8c6-e515-11e4-b510-962fcfab310\\_story.html](https://www.washingtonpost.com/local/crime/fbi-overstated-forensic-hair-matches-in-nearly-all-criminal-trials-for-decades/2015/04/18/39c8d8c6-e515-11e4-b510-962fcfab310_story.html) (discussing the FBI's announcement that the vast majority of its microscopic hair analysts gave flawed testimony over the two-decade period preceding 2000); Michael Shermer, *Can We Trust Crime Forensics?*, SCI. AM. (Sept. 1, 2015), <https://www.scientificamerican.com/article/can-we-trust-crime-forensics/> (discussing generally problems with forensic science).
- <sup>156</sup> For example, in an evidence ruling from the District of New Mexico, Judge William P. Taylor discussed the NRC Report and rulings from other district courts at length before ruling that, while the prosecution's ballistics expert could testify, he would not "be allowed to testify that he can conclude that there is a match to the exclusion, either practical or absolute, of all other guns." *United States v. Taylor*, 663 F. Supp. 2d 1170, 1178, 1180 (D.N.M. 2009). For other cases coming to the same conclusion with regard to ballistics evidence, see *United States v. Glynn*, 578 F. Supp. 2d 567, 570, 574-75 (S.D.N.Y. 2008), *United States v. Diaz*, No. CR 05-00167 WHA, 2007 WL 485967, at \*11-14 (N.D. Cal. Feb. 12, 2007), *United States v. Monteiro*, 407 F. Supp. 2d 351, 372-73 (D. Mass. 2006), and *United States v. Green*, 405 F. Supp. 2d 104, 108-09, 114-15 (D. Mass. 2005).
- <sup>157</sup> See generally DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011) (analyzing how and why humans rely on heuristics and biases to facilitate decisionmaking); Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CALIF. L. REV. 969 (2006) (analyzing the way humans rely on heuristics in the context of anti-discrimination law).
- <sup>158</sup> See generally Moriarty, *supra* note 146, at 310-11, 319-20 (arguing that judges unreasonably equate a long history of admission of evidence with reliability).
- <sup>159</sup> *United States v. Havvard*, 117 F. Supp. 2d 848, 854 (S.D. Ind. 2000).
- <sup>160</sup> See Moriarty, *supra* note 146, at 316 ("[A] long history of use confers no particular proof of validity."); see also Edwards & Mnookin, *supra* note 2 ("Respectfully, experience has shown that, at least in criminal trials, the suggestion that the 'adversarial system' represents an adequate means of demonstrating the unreliability of forensic evidence is mostly fanciful.").
- <sup>161</sup> Moriarty, *supra* note 146, at 316.
- <sup>162</sup> For example, cognitive scientists have recognized a heuristic known as the "halo effect." KAHNEMAN, *supra* note 157, at 82-85. The halo effect describes the human tendency to develop an initial positive or negative impression about someone, and then to fill in knowledge gaps or interpret new information in a way heavily colored by those initial assumptions. *Id.* For example, a professor who grades a single student's entire exam at once may find themselves heavily influenced in grading the second and third answers by how strong the answer to the first question was. *Id.* at 83-84.
- <sup>163</sup> ALLIANCE FOR JUSTICE, BROADENING THE BENCH: PROFESSIONAL DIVERSITY AND JUDICIAL NOMINATIONS 4, 6 (2016), <http://www.afj.org/wp-content/uploads/2014/11/Professional-Diversity-Report.pdf>. Among Obama's circuit court nominees, twenty-four had prosecutorial experience, while only five had experience in public defense. *Id.* at 11. Thirty-one circuit court nominees came from civil government backgrounds. *Id.* At the district court level, 126 nominees were prosecutors, while forty-five were public defenders. *Id.* at 12.
- <sup>164</sup> Gregory L. Acquaviva & John D. Castiglione, *Judicial Diversity on State Supreme Courts*, 39 SETON HALL L. REV. 1203, 1235 tbl.10 (2009).



- <sup>165</sup> See Walter W. Steele, Jr., *Unethical Prosecutors and Inadequate Discipline*, 38 SW. L.J. 965, 972 (1984) (expressing concerns about ethical violations under this organizational structure).
- <sup>166</sup> See, e.g., MATTHEW R. DUROSE & PATRICK A. LANGAN, BUREAU OF JUSTICE STATISTICS, FELONY SENTENCES IN STATE COURTS, 2002, at 6 (2004) (finding that while Black people made up twelve percent of the U.S. population, they comprised thirty-seven percent of the persons convicted of a violent felony); Bernadette Rabuy & Daniel Kopf, *Prisons of Poverty: Uncovering the Pre-Incarceration Incomes of the Imprisoned*, PRISON POLICY INITIATIVE (July 9, 2015), <https://www.prisonpolicy.org/reports/income.html> (finding that incarcerated people were earning forty-one percent less income prior to incarceration than non-incarcerated people of similar ages); see also KAREN DOLAN & JODI L. CARR, INST. FOR POLICY STUDIES, THE POOR GET PRISON: THE ALARMING SPREAD OF THE CRIMINALIZATION OF POVERTY (2015) (describing ways in which poverty leads to incarceration and vice versa).
- <sup>167</sup> Jolls & Sunstein, *supra* note 157, at 971. See generally John T. Jost et al., *The Existence of Implicit Bias Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies that No Manager Should Ignore*, 29 RES. ORGANIZATIONAL BEHAV. 39 (2009) (describing methods of studying implicit bias and summarizing prominent findings of implicit bias based on race and social status, among other criteria).
- <sup>168</sup> See Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 782-83 (2001) (observing that research has exposed cognitive biases among “doctors, real estate appraisers, engineers, accountants, options traders, military leaders, and psychologists”).
- <sup>169</sup> Commentators love to point this out. *E.g.*, *id.* at 784 (quoting JEROME FRANK, COURTS ON TRIAL: MYTH AND REALITY IN AMERICAN JUSTICE 410 (1949)).
- <sup>170</sup> See *id.* at 784, 816-17 (finding in a study of 167 federal magistrate judges that they were just as susceptible as lay people to anchoring, hindsight bias, and egocentric bias, but were less susceptible to framing and the representativeness heuristic).
- <sup>171</sup> See *id.* at 816-17 (discussing the influence of cognitive illusions on judges); Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1205, 1207, 1225 (2009) (finding in a study of 133 judges that implicit racial bias is “widespread”).
- <sup>172</sup> Chris Guthrie et al., *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1, 36 (2007) (proposing that judges delay making evidentiary rulings until the judge has had time to deliberate).
- <sup>173</sup> FED. R. EVID. 401.
- <sup>174</sup> FED. R. EVID. 401 advisory committee’s note on proposed rules (citing KENNETH S. BROUN ET AL., MCCORMICK ON EVIDENCE § 152 (7th ed. 2016)).
- <sup>175</sup> See, e.g., *United States v. Otero*, 849 F. Supp. 2d 425, 438 (D.N.J. 2012) (“[T]hrough the parties have not argued it, the relevance of [ballistics] testimony to the charges against Defendants is manifest. Clearly, the evidence will assist the trier of fact to determine a fact in issue ....”).
- <sup>176</sup> Jack B. Weinstein, *The Role of Judges in a Government of, by, and for the People: Notes for the Fifty-Eighth Cardozo Lecture*, 30

TRIAL JUDGES AND THE FORENSIC SCIENCE PROBLEM, 92 N.Y.U. L. Rev. 1532

CARDOZO L. REV. 1, 21 (2008).

177 United States v. Monteiro, 407 F. Supp. 2d 351, 358 (D. Mass. 2006).

178 **Daubert** v. Merrell Dow Pharm., Inc., 509 U.S. 579, 596 (1993).

179 People v. Kelly, 549 P.2d 1240, 1245 (Cal. 1976). Even in California, however, this rule comes with the caveat that “new evidence ... reflecting a change in the attitude of the scientific community” may allow courts to deviate from the precedential decision. *Id.*

180 *See, e.g.*, Moore v. Ashland Chem. Inc., 151 F.3d 269, 276 (5th Cir. 1998) (describing the **Daubert** standard as one which “requires some objective, independent validation of the expert’s methodology” and demands that the proponent “prove by a preponderance of the evidence that the testimony is reliable”); *Monteiro*, 407 F. Supp. 2d at 356 (rejecting the government’s argument that “because toolmark identification evidence has been deemed admissible by many other courts, the burden of proving such evidence to be unreliable should shift to the defendants”); Eaton Corp. v. Parker-Hannifin Corp., 292 F. Supp. 2d 555, 567 (D. Del. 2003) (“[T]he proponent of the expert testimony ... must prove by a preponderance of the evidence that the testimony is reliable.”).

181 *See* United States v. Hebshie, 754 F. Supp. 2d 89, 114-15 (D. Mass. 2010) (suggesting that arson analysis is “bad science” and emphasizing that the admissibility standard for evidence in criminal cases is identical to the civil standard); Gertner, *supra* note 2 (criticizing shoddy forensic evidence and judges who uncritically admit it).

182 405 F. Supp. 2d 104 (D. Mass. 2005).

183 *Id.* at 122.

184 *Id.* at 123.

185 *Id.* Judge Gertner admitted the ballistics testimony in *Green*, albeit cabining the expert to “testify[ing] to his observations” without “conclud[ing] that the match he found by dint of the specific methodology he used permits ‘the exclusion of all other guns’ as the source of the shell casings.” *Id.* at 124. If a judge is actually considering departing from the routine practice of her jurisdiction or the judiciary as a whole, there is yet another incentive to toe the party line: the prospect of undermining previous convictions. This concern may be animated by specific cases over which the judge in question presided in the past, or it may be a more general concern about the convictions secured on the basis of a certain type of forensic testimony in general. Thus, a judge faced with Defendant F may have at the back of her mind the past convictions of Defendants A-E when she rules to admit evidence that is increasingly being called into question outside the courtroom.

186 *See, e.g.*, KATE BERRY, BRENNAN CENTER FOR JUSTICE, HOW JUDICIAL ELECTIONS IMPACT CRIMINAL CASES 4 (2015) (describing one attack ad that targeted former Kentucky Supreme Court Justice Janet Stumbo for having “sided with criminals 50 percent of the time” and noting that Stumbo ultimately lost the election); *see also* Liz Seaton, *Smear Campaigns Against Justice*, U.S. NEWS (Feb. 29, 2016, 8:00 AM), <https://www.usnews.com/opinion/articles/2016-02-29/supreme-court-sees-how-judicial-elections-mar-criminal-justice-reform> (finding that more than half of the air time for judicial election advertising was focused on tough-on-crime messages).

187 Joanna Cohn Weiss, Note, *Tough on Crime: How Campaigns for State Judiciary Violate Criminal Defendants’ Due Process Rights*, 81 N.Y.U. L. REV. 1101, 1106 (2006) (quoting Mark Hansen, *When Is Speech Too Free?*, A.B.A. J., May 2001, at 20 (discussing a 2011 California judicial election)).

- <sup>188</sup> See, e.g., BERRY, *supra* note 186, at 5 (describing an attack ad in which a challenger alleged that the incumbent had released a convicted child rapist who then continued to commit offenses).
- <sup>189</sup> *Id.* at 5 (posting a video of *Tough on Predators* (Law Enforcement Alliance of Am. commercial broadcast May 9, 2014), <http://newpoliticsreport.org/spot/leaa-tough-on-predators-2/>); see also Weiss, *supra* note 187, at 1105 (describing an attack campaign from the 1994 race for the Texas Court of Criminal Appeals in which “one candidate and former prosecutor called the Court of Criminal Appeals a ‘citadel of technicality’” (internal citations omitted)).
- <sup>190</sup> BERRY, *supra* note 186, at 1 (finding this result consistently across ten “recent, prominent, and widely cited empirical studies” on the subject).
- <sup>191</sup> *Id.* at 8-9 (discussing the findings of a 2014 American Constitution Society study).
- <sup>192</sup> FED. R. EVID. 702.
- <sup>193</sup> See *supra* note 180 (collecting sources describing Rule 104’s requirements).
- <sup>194</sup> See *supra* note 156 (collecting cases where testimony was cabined in the context of ballistics evidence).
- <sup>195</sup> See *United States v. Monteiro*, 407 F. Supp. 2d 351, 356 (D. Mass. 2006) (rejecting the government’s argument that “because toolmark identification evidence has been deemed admissible by many other courts, the burden of proving such evidence to be unreliable should shift to the defendants”).
- <sup>196</sup> See PCAST REPORT, *supra* note 1, at 84, 87 (concluding that additional research into bite-mark analysis is inadvisable because “the prospects of developing bitemark analysis into a scientifically valid method [are] low”).
- <sup>197</sup> See, e.g., Microscopic Hair Analysis Press Release, *supra* note 4 (reporting widespread false testimony in the realm of microscopic hair analysis); Press Release, FBI, FBI Laboratory Announces Discontinuation of Bullet Lead Examinations (Sept. 1, 2005), <https://archives.fbi.gov/archives/news/pressrel/press-releases/fbi-laboratory-announces-discontinuation-of-bullet-lead-examinations> [hereinafter CBLA Press Release] (“While the FBI Laboratory still firmly supports the scientific foundation of bullet lead analysis, given the costs of maintaining the equipment, the resources necessary to do the examination, and its relative probative value, the FBI Laboratory has decided that it will no longer conduct this exam.”).
- <sup>198</sup> See, e.g., PCAST REPORT, *supra* note 1, at 82 (recommending further study of DNA analysis of complex-mixture samples); *id.* at 96-97 (concluding that “only two properly designed studies of the accuracy of latent fingerprint analysis have been conducted” and recommending further study); *id.* at 111 (finding that “only a single appropriately designed study” of firearms analysis has been conducted, and “[t]here is thus a need for additional, appropriately designed black-box studies to provide estimates of reliability”); *id.* at 117 (concluding that the validity of footwear analysis is “unsupported by any meaningful evidence” and is therefore “not scientifically valid”).
- <sup>199</sup> Cf. KAHNEMAN, *supra* note 157, at 85-88 (describing the cognitive illusion that “what you see is all there is,” meaning that people are eager to draw conclusions based on the information available and struggle to suspend judgment even when it is clear that they are lacking all relevant information).

- 200 See Paul C. Giannelli, *Comparative Bullet Lead Analysis: A Retrospective*, 47 CRIM. L. BULL. 306, 306 (2011) (“For over thirty years, FBI experts testified about [CBLA] ....”).
- 201 See CBLA Press Release, *supra* note 197 (“Bullet lead examinations have historically been performed in limited circumstances, typically when a firearm has not been recovered or when a fired bullet is too mutilated for comparison of physical markings.”).
- 202 Giannelli, *supra* note 200, at 306.
- 203 NAT’L RESEARCH COUNCIL, FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE 112-13 (2004) [hereinafter CBLA REPORT] (concluding that because “[d]etailed patterns of distribution of ammunition are unknown ... an expert should not testify as to the probability that a crime scene bullet came from the defendant”); *id.* at 113 (concluding that because “[t]he available data do not support any statement that a crime bullet came from, or is likely to have come from, a particular box of ammunition” testimony should not assert that “the crime bullet came from the defendant’s box or from a box manufactured at the same time”). Microscopic hair analysis suffers from the same shortcoming, namely that the commonality of hair characteristics in the general population is unknown. See *supra* notes 44-45 and accompanying text.
- 204 United States v. Mikos, No. 02 CR 137, 2003 WL 22922197 (N.D. Ill. Dec. 9, 2003).
- 205 *Id.* at \*1 (“[T]he defense claims that the Government has supplied no underlying basis or support for [the expert’s] conclusion that, because these lead samples are ‘analytically indistinguishable,’ the bullets from the victim and the bullets from the cartridges in the box ‘likely originated from the same manufacturers’ source (melt) of lead.”).
- 206 *Id.* at \*6.
- 207 *Id.* at \*5.
- 208 *Id.* A similar argument could be made for the assertion that all human mouths are unique. See *supra* note 34 and accompanying text.
- 209 Mikos, 2003 WL 22922197, at \*6.
- 210 *Id.*
- 211 CBLA REPORT, *supra* note 203, at 90-94, 106-08.
- 212 See CBLA Press Release, *supra* note 197 (announcing that the FBI “will no longer conduct the examination of bullet lead,” largely because “neither scientists nor bullet manufacturers are able to definitely attest to the significance of an association made between bullets in the course of a bullet lead examination”).
- 213 CBLA REPORT, *supra* note 203, at ix.
- 214 See Giannelli, *supra* note 200, at 310 (noting that Mikos was the first case to rule such evidence inadmissible).





**2010 Utah L. Rev. 299**

Utah Law Review  
2010

Symposium

Lessons from the Lab: Implications of the 2009 National Academy of Sciences Report on the Future of Forensic Science

Jane Campbell Moriarty<sup>at</sup>

Copyright (c) 2010 Utah Law Review Society; Jane Campbell Moriarty

## **WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON FORENSIC SCIENCE AND THE ROLE OF THE JUDICIARY**

1993

[The trial judge should undertake] a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue. We are confident that federal judges possess the capacity to undertake this review.<sup>1</sup>

2009

With the exception of nuclear DNA analysis . . . no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty . . . demonstrate a connection between evidence and a specific individual or source.<sup>2</sup>

“[T]he undeniable reality is that the community of forensic science professionals has not done nearly as much as it reasonably could have done to establish either the validity of its approach or the accuracy of its practitioners’ conclusions,”<sup>3</sup> and the courts have been “utterly ineffective” in addressing this problem.<sup>4</sup>

### **I. Introduction**

For several decades, the prosecution and its witnesses have maintained that despite little research and virtually no standards, they can match a fingerprint, \*300 handwriting, bullet and bullet cartridge, hair, dental imprint, footprint, tire track, or even a lip print to its unique source (collectively, “individualization evidence”). Not only can they match it, they claim, they can do so often without any error rate.

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

In the last few decades, with the help of lawyers and academics, litigants have challenged the underlying reliability of individualization evidence. Scholars in various disciplines have written about the startling state of individualization evidence, including its lack of standards, research, and established error rates, and its failure to rely upon statistical probabilities to estimate the likelihood of a match. Since its inception, the Innocence Project has exonerated more than 250 people, a majority of whose convictions have involved inaccurate or even fraudulent forensic science testimony, including individualization evidence.

Despite the lack of proof that such evidence is scientifically reliable (and continued exculpations), courts have rejected most challenges to individualization evidence and continue to admit such testimony. With every exoneration, proof mounts that forensic science cannot do what it claims to be able to do with the precision alleged. By not requiring minimal standards for the reliability of individualization evidence, courts have allowed the forensic science system to operate without any checks and balances and to convict innocent people in numbers we can only estimate.

In February 2009, the National Academy of Sciences issued its long-awaited and groundbreaking report on the status of forensic science, *Strengthening Forensic Science in the United States: A Path Forward* (“the NAS Report”).<sup>5</sup> The NAS Report is a scathing indictment of both the state of the forensic science system and judicial rulings on such individualization evidence.

This Article discusses the findings of the NAS Report, relevant cases that predate the report, and some cases decided since the report. It posits that the judiciary, which has created a standard of reliability, has failed to hold prosecutorial expert evidence to that standard. Using examples from history and modern cognitive science explanations, the Article tries to explain why the judiciary has been so unwilling to rigorously examine forensic science evidence and urges the judiciary to rethink its perspective going forward.

While the NAS Report suggests an overhaul of the current system, that overhaul is a contentious idea that may well not occur in the near (or even longer) future. Thus, a current crisis exists that the judiciary must address in its day-to-day decision making. The Article suggests how the judiciary can become a more effective crucible for testing the strength and limitations of forensic science.

## II. The NAS Report

In February 2009, the National Academy of Sciences issued its report on the status of forensic science, *Strengthening Forensic Science in the United States: A Path Forward*.<sup>6</sup> It is a detailed discussion, with each chapter providing a \*301 compendium of a separate subject relevant to forensic science. The NAS Report, highlighting the myriad shortcomings and failures of what we call forensic science, seems quite shocked at the current situation, remarking that “[t]he Law’s greatest dilemma in its heavy reliance on forensic evidence, however, concerns the question of whether--and to what extent--there is any science in any given ‘forensic science’ discipline.”<sup>7</sup>

In addition to detailing the shortcomings of the scientific evidence, the NAS Report explains how the judicial system has utterly failed in its regulation of forensic science in criminal cases.<sup>8</sup> Due to the system failure on every level to improve the quality of forensic science, the report calls for a virtually complete overhaul of the forensic science system, suggesting the creation of the National Institute of Forensic Science (NIFS), an independent agency to oversee forensic science in the United States.<sup>9</sup>

By any interpretation, the report is a critique of both the current state of forensics as practiced in the United States and the judiciary’s unwillingness or inability to require minimum standards for forensic evidence. While the NAS Report compliments the forensic science community on the “valuable evidence that has contributed to the successful prosecution and conviction of criminals as well as to the exoneration of innocent people,”<sup>10</sup> it simultaneously cautions that, although forensic science has advanced, it is now clear that “in some cases, substantive information and testimony based on faulty forensic science analyses may have contributed to wrongful convictions.”<sup>11</sup> These wrongful convictions, the NAS Report notes, reflect

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

the “potential danger of giving undue weight to evidence and testimony derived from imperfect testing and analysis.”<sup>12</sup> Additionally, “imprecise or exaggerated expert testimony has sometimes contributed to the admission of erroneous or misleading evidence.”<sup>13</sup>

The reasons for the problems inherent in forensic science in the United States, the report explains, are myriad, including the following:

- Disparities in the forensic science system between federal resources and states and in the various standards for the medical examination system;<sup>14</sup>
- Lack of mandatory standardization, certification, and accreditation of laboratories, as well as no uniformity in the certification of forensic practitioners;<sup>15</sup>
- **\*302** · A wide variability across disciplines with respect to techniques, methodologies, reliability, types and numbers of potential errors, research, general acceptability, and published material;<sup>16</sup> and
- Problems relating to interpretation of forensic science--the questionable conclusion about individualization where a specimen is “matched” to a particular individual or source.<sup>17</sup>

The NAS Report emphasizes that “law enforcement officials and the members of society they serve need to be assured that forensic techniques are reliable.”<sup>18</sup> To that end, the NAS Report calls for a body of research to establish the limits and measures of performance of the various forensic disciplines, as well as research to address the impact of sources of variability and potential bias.<sup>19</sup> Without this research, the NAS Report insists, “[w]e must limit the risk of having the reliability of certain forensic science methodologies judicially certified before the techniques have been properly studied and their accuracy verified by the forensic science community.”<sup>20</sup> Nonetheless, the NAS Report concedes that “some courts appear to be loath to insist on such research as a condition of admitting forensic science evidence in criminal cases,”<sup>21</sup> apparently believing the forensic science disciplines are currently incapable of offering such validation.<sup>22</sup> The NAS Report finds that the judicial approach to forensic disciplines has been “if you can’t meet the standard, we’ll eliminate the standard”--a frightening approach, given the clear concordance between forensic science and actual guilt or innocence.<sup>23</sup>

In addition to arson,<sup>24</sup> biological evidence, chemical analysis, and the medical examination system in the United States, the NAS Report reviews non-DNA individualization evidence (fingerprints, hair, handwriting, toolmarks, shoeprints and tire tracks, and forensic odontology, among others).<sup>25</sup> The report describes what a conclusion of individualization actually requires:

- **\*303** [A] conclusion of individualization implies that the evidence originated from that source, to the exclusion of all possible sources. The determination of uniqueness requires measurement of object attributes, data collected on the population frequency of variation in these attributes, testing of attribute independence, and calculation of the probability that different objects share a common set of observable attributes.<sup>26</sup>

DNA evidence, mentioned favorably in the report, possesses all these attributes, which are hallmarks of a proper conclusion about the probability of a match.<sup>27</sup> By contrast, the remaining categories of individualization evidence, including fingerprint comparison, possess none.<sup>28</sup> The report concludes that “no forensic method other than nuclear DNA analysis has been rigorously shown to have the capacity to consistently and with a high degree of certainty support conclusions about . . . ‘matching’ of an unknown item of evidence to a specific known source.”<sup>29</sup> The report also notes that the forensic science community has had “little opportunity to pursue or become proficient in the research that is needed to support what it does.”<sup>30</sup>

Despite the need for such research to be done to validate methodology and underlying presumptions, the prosecution continues to argue such evidence is reliable, and, to date, many courts seem to be agreeing.<sup>31</sup> When describing the lack of



research or scientific scrutiny concerning individualization specialties, the report concludes that “although the precise error rates of these forensic tests are still unknown, comparison of their results with DNA testing in the same cases has revealed that some of these analyses, as currently performed, produce erroneous results.”<sup>32</sup> These “erroneous results” caused wrongful convictions in some cases. Although people on both sides of the aisle may debate how important this lack of \*304 research is, the inescapable conclusion is that without scientific proof of the foundations necessary to legitimately declare a match, much expert testimony is simply a hunch, supported by experiential “observations of countless samples.”<sup>33</sup> There are numerous reasons that judges admit such testimony so readily: they trust the FBI and other forensic scientists; they assume they would have heard on a more global level if forensic science were unreliable; they share an intuitive belief that forensic comparison is valid; they are comforted by other courts’ decisions admitting the evidence (what I term the “generally accepted by other courts” standard);<sup>34</sup> they find the analysts’ experiential knowledge convincing; and they believe that the long history of use confers some field reliability to the work.<sup>35</sup> Although many of these justifications might seem compelling on an intuitive, “hunch-like” level, further scrutiny reveals some serious cracks in the foundation of such opinions.<sup>36</sup>

### \*305 III. The Judicial Role in the Gatekeeping of Individualization Evidence

During his confirmation hearing in 2005, then-Judge John Roberts told Congress, “Judges are like umpires. Umpires don’t make the rules; they apply them. The role of an umpire and a judge is critical. They make sure everybody plays by the rules, but it is a limited role.”<sup>37</sup> There is some nugget of truth in this partial description of what judges do, although there is long-standing, serious disagreement that judges do not “make law.”<sup>38</sup> The comment, however, seems disingenuous in its overly facile, folksy explanation of the role of the judiciary. And not surprisingly, Justice Roberts’s comment sparked much controversy, not the least of which was from Judge Richard Posner, who sniffed dismissively that “[n]o serious person thinks that the rules that judges in our system apply, particularly appellate judges and most particularly the Justices of the U.S. Supreme Court, are given to them the way the rules of baseball are given to umpires. The rules are created by the judges themselves.”<sup>39</sup>

Proving the accuracy of Judge Posner’s comments, the Supreme Court has created, from virtual whole cloth, the “rules” governing admissibility of expert testimony, in the Daubert/Joiner/Kumho Tire trilogy (“trilogy”).<sup>40</sup> To be admissible, the party seeking to introduce forensic evidence must be able to establish, to the satisfaction of the trial court, that the proposed expert evidence meets the Supreme Court-created standard of reliability, incorporated into the Federal Rules of Evidence in 2000.<sup>41</sup> A substantial minority of courts use versions \*306 of the so-called “Frye general acceptance test,” while the federal courts and a majority of state jurisdictions use some variant of the Daubert reliability standard.<sup>42</sup>

Collectively, these cases (and the amended Federal Rule of Evidence 702) reflect a concern about whether evidence is reliable as used in a given case. “[R]eliability cannot be judged globally, ‘as drafted,’ but only specifically, ‘as applied.’ The emphasis [is] on the judgment of reliability as it applies to the individual case, to the ‘task at hand.’”<sup>43</sup>

The trilogy governing the admission of expert testimony claims to envision a flexible standard, in which the trial court, as gatekeeper of the evidence, determines whether expert evidence meets a minimal standard of evidentiary reliability.<sup>44</sup> Thus, in the case of expert evidence in the federal courts, the Supreme Court has created the rules by which the courts and litigants must abide<sup>45</sup> and requires the inferior courts to serve as the arbiters of whether the evidence \*307 complies with those rules in this judicially-created gatekeeping role.<sup>46</sup> Thus, the judiciary has created an entire universe where none previously existed and has appointed itself the master of such universe.

It seems beyond cavil to hope judges would make sure “everybody plays by the rules” the Supreme Court and other federal courts have taken great pains (and several years) to create, refine, and develop. At a minimum, we expect our judiciary to attempt to level the playing field for all participants and to apply equal standards to competing litigants.

Nonetheless, after reading a multitude of cases involving forensic science evidence and empirically driven studies about what courts have done post-trilogy, it is not at all clear that judges “make sure everybody plays by the rules.” Indeed, research suggests that many judges do not require the prosecution to play by the same rules that other litigants in both civil and

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

criminal cases do when the subject is forensic science evidence. These findings are not based solely upon my own impressions,<sup>47</sup> but rather upon empirical data collected and analyzed by others<sup>48</sup> and the conclusions of the NAS Report: “[T]he vast majority of the reported opinions in criminal cases indicate that trial judges rarely exclude or restrict expert testimony offered by prosecutors; most reported opinions also indicate that appellate courts routinely deny appeals contesting trial court decisions admitting forensic evidence against criminal defendants.”<sup>49</sup>

It seems abundantly clear that many courts merely give lip service to pressing the prosecution to meet its burden of proof in Daubert hearings; much the way some bartenders mix an extra dry martini, as if waving the bottle of vermouth next to the glass will suffice. In other words, the court will find in favor of the \*308 government’s evidence as long as the government makes the appropriate motions.<sup>50</sup>

Deirdre Dwyer writes compellingly about the asymmetries in the application of Daubert between civil and criminal cases.<sup>51</sup> She considers the normative expectations that either all evidence would be treated equally, or that evidence rules “in a criminal action [would be] geared more towards reducing the risk of an erroneous outcome than they are in a civil action.”<sup>52</sup> After looking at data-driven studies, however, she concludes, “[i]t would seem that the expert evidence of civil plaintiffs, particularly in toxic tort cases, is subject to greater scrutiny than that of civil defendants, while the expert evidence of criminal prosecutors is subject to less scrutiny than that of criminal defendants, or than that of civil parties.”<sup>53</sup> This seems to be an odd outcome, indeed, when civil cases involve only monetary damages, while criminal cases deal with stakes of much greater value.<sup>54</sup>

Courts, including the U.S. Supreme Court, have been willing to engage in detailed, scientific analysis of proposed expert evidence in civil cases.<sup>55</sup> In many of \*309 these cases, the district courts, affirmed by courts of appeal on an abuse of discretion standard, have found such testimony insufficiently reliable.<sup>56</sup> Indeed, the NAS Report notes the different standards in civil and criminal cases, remarking that “courts have not . . . imposed standards ensuring the application of scientifically valid reasoning and reliable methodology in criminal cases involving Daubert questions.”<sup>57</sup> Furthermore, the report concludes that upon reviewing the reported decisions, “at least in criminal cases, forensic science evidence is not routinely scrutinized pursuant to the standard of reliability enunciated in Daubert.”<sup>58</sup>

Since *Daubert v. Merrell Dow Pharmaceuticals*, there have been numerous, substantial challenges to the reliability of many types of forensic science, particularly in the area of individualization evidence.<sup>59</sup> Federal courts have rejected most defense challenges made to the reliability of fingerprint individualization evidence.<sup>60</sup> With a handful of exceptions in which trials courts have limited conclusions about handwriting and ballistics,<sup>61</sup> courts have rejected challenges to \*310 the reliability of those forms of evidence as well.<sup>62</sup> Microscopic hair comparison, implicated for its role in wrongful convictions in many exonerations,<sup>63</sup> has been admitted by both federal and state courts, with seemingly no concern for the potentially substantial rate of error such evidence presents.<sup>64</sup>

Courts do not admit such evidence because it meets the trilogy and FRE 702’s standards of reliability. To the contrary, prosecutors and their experts cannot establish the validity of what they claim to be able to do with the precision alleged--a point made quite clear in the NAS Report where it remarks that there is a “notable dearth” of scientific studies to establish the validity of many forensic science methods.<sup>65</sup> Rather, courts admit such evidence simply because they cannot seem to imagine doing otherwise. “The methods of latent print identification can and have been tested. They have been tested for roughly 100 years. They have \*311 been tested in adversarial proceedings with the highest possible stakes--liberty and sometimes life.”<sup>66</sup>

In *United States v. Llera Plaza*, Judge Pollack vacated his earlier opinion limiting fingerprint examiners’ testimony to pointing out similarities (rather than testifying as to the conclusion of a match) and decided it was appropriate to permit testimony about conclusions.<sup>67</sup> “In short,” he begins, “I have changed my mind.”<sup>68</sup> Relying in part upon testimony admitted in the rehearing he granted the government, he concludes if it is “sufficiently reliable in England,” it should be good enough for U.S. courts.<sup>69</sup> The *Llera Plaza* court, like many others, was impelled by the long history of fingerprint admission, noting that both English and American courts have admitted fingerprint comparison testimony for nearly a century.<sup>70</sup>

In *United States v. Prime*, the district court stated when allowing conclusions about handwriting, “[w]here a technique has

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

been repeatedly applied and tested by law enforcement and courts for over a century, the court does not believe the absence of scientific data, without more, should sound the death knell for such testimony.<sup>71</sup> The Court of Appeals for the Ninth Circuit upheld this decision, deciding that as long as the process is generally reliable, any potential error can be brought to the jury's attention with cross-examination and competing testimony.<sup>72</sup> While the language the court used is not unusual, what the court found to be "generally reliable" was based upon admitted evidence that the error rate for handwriting comparison was approximately 13 percent!<sup>73</sup> Yet the court was not convinced that such evidence might not meet the standard of reliability. One must wonder then at what rate of error does the evidence become unreliable?

A most unusual form of individualization evidence was admitted in *State v. Davis*, where the court allowed testimony that the defendant's lip print matched a \*312 lip print at the scene.<sup>74</sup> Although there was absolutely no foundation for such evidence, the trial court, without the smallest amount of critical inquiry, found such proof met both Frye and reliability standards.<sup>75</sup> The defense lawyer handling this case acted incompetently by not satisfactorily challenging the testimony,<sup>76</sup> yet both the trial court and the court of appeals in the direct appeal were satisfied with little more than the ipse dixit testimony of the government's witnesses.<sup>77</sup>

While this lip print case may be an outlier in terms of the courts' willingness to accept expert evidence in criminal cases without sufficient foundational proof of reliability (or, indeed, even general acceptance), it is in many ways emblematic in its willingness to take prosecutorial experts at their word with no requirement of supporting data.<sup>78</sup>

Bite mark evidence, also known by its more technical name "forensic odontology," has a history of acceptance by courts despite the fact that the science supporting it is dubious. In fact, the theories upon which the field is predicated--that dentition is unique and that marks found upon a victim can be linked unequivocally to the perpetrator--are not supported by current data.<sup>79</sup> The problems, like the rest of forensic science individualization evidence, are legion. It is unlikely that human dentition is unique; very few teeth are actually used to make a bite mark; and bite marks become distorted on human skin due to the passage of time, thus not maintaining the accuracy of the marks.<sup>80</sup> Moreover, unlike DNA evidence, which uses databases to generate probabilities that the suspect left his DNA at the crime scene, there is no bite mark database from which probabilities can be generated<sup>81</sup>-- although such probabilistic testimony has been admitted \*313 (without accurate foundation) in trials.<sup>82</sup> Bite mark analysts often testify using phrases such as "consistent with" or "positively match."<sup>83</sup> Yet such statements have no grounding in science, because they are purely subjective conclusions unsupported by data.<sup>84</sup> In addition, as with other forms of individualization forensic science evidence, statements about a match are neither scientifically supportable<sup>85</sup> nor scientifically meaningful.<sup>86</sup> Early research, however, appears to indicate that such statements are convincing to juries.<sup>87</sup>

The NAS Report finds plenty to criticize about forensic odontology, concluding that no scientific studies support the claims that odontologists can demonstrate sufficient detail for positive identification. In numerous instances, "experts diverge widely in their evaluations of the same bite mark evidence," which raises serious questions about the value and scientific objectivity of the discipline.<sup>88</sup> Of course, like every other form of individualization evidence, judges have admitted it readily into the courtroom.<sup>89</sup>

In *United States v. Crisp*, the court admitted both fingerprint comparison and handwriting comparison evidence despite strong, solid challenges to the evidence.<sup>90</sup> Regarding fingerprint comparison, the court held that "the principles underlying fingerprint identification . . . bear the imprimatur of a strong general acceptance, not only in the expert community, but in the courts as well. . . . Put simply, . . . [there is] no reason . . . to believe that this general acceptance of the \*314 principles underlying fingerprint identification has, for decades, been misplaced."<sup>91</sup> Thus, in a sweeping but unsupported statement, the court concluded that "the district court was well within its discretion in accepting at face value the consensus of the expert and judicial communities that the fingerprint identification technique is reliable."<sup>92</sup> With regard to handwriting comparison, the court stated, "[t]he fact that handwriting comparison analysis has achieved widespread and lasting acceptance in the expert community gives us the assurance of reliability that Daubert requires."<sup>93</sup> By not actually evaluating the testimony using Daubert's suggested factors,<sup>94</sup> the court evaded the problem that the evidence did not meet reliability requirements, a point noted by the dissent, which explained in detail how the government "utterly failed to meet its burden" of establishing reliability under Daubert.<sup>95</sup>

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

The NAS Report criticizes both handwriting and fingerprint comparison. The report finds that the standards for fingerprint comparison are subjective, as are the declarations of a match.<sup>96</sup> In fact, not only is the outcome of analysis not necessarily repeatable from one examiner to another, research cited by the report indicates that experienced examiners do not even necessarily agree with their own prior conclusions!<sup>97</sup> Analysis of the methodology used by the government to declare fingerprint matches is so entirely without foundation or objective standards that even the validity of the method cannot be tested.<sup>98</sup> Moreover, even if the two foundations of the specialty are true--that fingerprints are unique and persist unchanged throughout life--that does not imply that “anyone can reliably discern whether or not two friction ridge impressions were made by the same person.”<sup>99</sup> Similarly, handwriting comparison fares even more poorly. Conclusions are entirely subjective, and the NAS Report recommends that its “scientific basis . . . needs to be strengthened.”<sup>100</sup>

Regarding microscopic hair comparison, the NAS Report was even more critical, commenting on its high rate of error and stating “the committee found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA.”<sup>101</sup> In sum, the non-DNA individualization evidence is \*315 simply unsupported by research, and serious questions are raised about its evidentiary reliability.

Yet the report is willing to absolve the courts of any real responsibility in enforcing the reliability mandate, citing various reasons: judicial decisions about reliability are “flexible,” and much discretion is granted to the trial court;<sup>102</sup> judges (and lawyers) “lack the scientific expertise necessary to comprehend and evaluate forensic science in an informed manner”;<sup>103</sup> judges work alone and “often [have] little time for extensive research and reflection”; and, echoing many others, the adversarial process is “not suited to the task of finding ‘scientific truth.’”<sup>104</sup>

There is merit in each of these justifications for courts’ unwillingness to subject most of forensic science to any meaningful analysis. And yet, the justifications seem weak, particularly in light of those same courts’ ability to tackle far more complicated questions of scientific causation in exceptionally technical civil cases. And the justification is sorely inadequate when considering the difference in what is at stake in the respective cases. Regarding challenges to fingerprint comparison, Professor Mnookin writes, “The real embarrassment is the way that the courts have been a willing accomplice in this turf battle, in the process abnegating their gatekeeping responsibilities under Daubert.”<sup>105</sup>

In civil cases, courts seem quite up to the task of evaluating microbiology, teratology, and toxicology evidence, discussing both science and statistics with plenty of acumen.<sup>106</sup> Yet when it comes to evaluating the shortcomings of lip prints and handwriting, courts are unable to muster the most minimal grasp of why a standardless form of comparison might lack evidentiary reliability or trustworthiness.<sup>107</sup> Rather, they have relied on tired aphorisms and biased heuristics so as to avoid stating the obvious: there are no standards, there is no legitimate basis for the matches testified to, and forensic individualization is currently built on sand.

**\*316 IV. “History May Be Servitude”:<sup>108</sup> The Erroneous Judicial Focus on the Long History of Use of Individualization Evidence**

Courts have found a variety of reasons to admit individualization evidence, yet one of the most common justifications courts provide is the long history of use. Yet a long history of use confers no particular proof of validity. From the time of Nicader of Colophain, 200 B.C., through the height of their popularity in the mid-nineteenth century, leeches were used to “bleed” the sick as a form of treatment for virtually all ailments, from pneumonia to hemorrhoids.<sup>109</sup> As late as 1920, the use of lancets and leeches for bloodletting was favored by some physicians to treat pneumonia.<sup>110</sup> One would think that reflective consideration would have shown the lack of bloodletting efficacy, yet physicians continued to employ it, year after year, decade after decade, century after century. It was not until the rise of experimental methodology and restraints on empirical methods began to gain ground that the use of bloodletting and leeches in everyday medicine began to drop off and subsequently evaporate.<sup>111</sup>

Likewise, the “long history of use” argument failed to establish underlying validity of expert testimony in the Salem witchcraft trials. Cotton Mather and those overseeing the trials approved of the well-known publications<sup>112</sup> providing expert



WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

guidance for courts on the proof of witchcraft--expert treatises that had stood the test of time in other prosecutions.<sup>113</sup> The principles of the "experts" were \*317 recognized by courts for nearly a century, which was a major reason the courts in Salem were swayed by such opinion.

The Oracle of Delphi provides yet a third example of the fallacy of historical reliance.<sup>114</sup> Citizens, both of high and low status, would approach the oracle and ask questions, often about quotidian matters such as whether to marry or matters of a more serious nature, such as whether to resist the Persian invasion.<sup>115</sup> Despite advice from the oracle not always proving prescient or wise,<sup>116</sup> the oracle was consulted for over a thousand years.<sup>117</sup>

We now appreciate that oracles cannot predict future events, that witchcraft is not real, and that bleeding is generally not effective to cure disease, but old ideas die hard. Perhaps there is a lesson here useful to the judiciary.

### V. Judicial Decision Making and the Problems of Cognitive Bias

Legal, psychological, and sociological scholars have all examined and opined about judicial decision making to determine how judges decide cases.<sup>118</sup> The methods of analysis and theories posed are varied, rich, and complex, suggesting that decision making is a product of reason and intuition. Some find that political agendas or background and experience inform decision making, while others argue that judges are influenced by precedent. One theme, however, that resonates throughout much of the literature is that "judges are human."<sup>119</sup> They are swayed \*318 by heuristic decision making, friendships, beauty, the strength of a case, public opinion, fear of reversal, and the normal set of cognitive biases to which we all are subject: expectation bias, hindsight bias, confirmation bias, tunnel vision, and so forth.

In an interesting empirical article, *Blinking on the Bench: How Judges Decide Cases*, the authors write about the dual-process models of cognition (intuitive and deliberative).<sup>120</sup> Roughly explained, intuitive decision making is spontaneous, effortless, relies on pre-existing heuristics, and is fast,<sup>121</sup> while the deliberative process is slower, requiring more "effort, motivation, concentration, and the execution of learned rules."<sup>122</sup> The *Blinking on the Bench* authors conclude that judges tend to use intuitive decision making for the everyday problems they see on the bench.<sup>123</sup> Thus, "[w]hen ruling on the admissibility of evidence at trial, judges often have little choice but to think intuitively."<sup>124</sup>

Many might consider intuitive decision making in this role to be beneficial because judges have "typical" evidence questions, think about admissibility questions all the time, and have abundant experiential knowledge that informs the snap judgments that are needed at trial. The use of heuristics to think quickly and decide intuitively has great value. We don't deliberate when a ball is thrown to us; we either catch it or get bonked on the head. Similarly, in trial, judges need to decide evidentiary questions in real time. They cannot over-deliberate every time an objection is raised.

But there is a decided downside to such quick thinking. There is growing evidence that intuitive and impressionistic decisions about evidence may be more error-prone than a more deliberative process. In an experiment with judges, researchers discovered that when subjects were given a problem they thought they \*319 could solve intuitively, the error rate was substantially higher than with a problem the judges had to actually reason their way through to resolve.<sup>125</sup>

Moreover, the problem of cognitive bias is most apparent, according to researchers, in the use of intuitive decision making, and it is where such problems as stereotyping, prejudice, and discrimination are likely to arise.<sup>126</sup> Relying on heuristic shortcomings can lead to systematic biases.<sup>127</sup>

Thus, it may be worthwhile to consider both the dual processing systems and bias when thinking about courts' decisions on individualizing evidence in the past and going forward into the future.

Part of the courts' persistence in finding individualization specialties reliable may be due to judges using intuitive decision making both before trial and during trial when resolving admissibility questions about fingerprinting, hair, and handwriting comparison. As noted in *Blinking on the Bench*, when judges are presented with a problem they perceive to be simple, they

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

use heuristics to solve it--potentially making errors and not recognizing them due to the belief that the issue before them is not complex. It may well be that fingerprint, handwriting, and microscopic hair comparisons seem to be "simple" problems for judges, leading them to use intuitive decision making about the outcome (although with some rational processing, of course, in the written decision).

Consider the diametric approaches courts employ when deciding admissibility of scientifically complex expert evidence (such as DNA comparison or toxic tort causation) versus the non-DNA individualization evidence such as fingerprints and handwriting. In the complex scientific evidence cases, courts appear to use a more rational processing system and engage in deliberate, analytic reasoning throughout the opinion to determine whether the evidence is reliable.<sup>128</sup> A good example of this methodology is found in the Supreme Court's opinion in *General Electric v. Joiner*, in which the court engaged in a long methodological analysis of the quality of epidemiological evidence to determine whether plaintiff's proof that PCB exposure had promoted his lung cancer was sufficiently reliable to be admitted.<sup>129</sup> None of the writing appears intuitively-based.

By contrast, very few cases involving individualization evidence seem to be of the long, methodological analysis present in *Joiner*. To the contrary, courts seem to rely on such heuristic devices as "long history of use" or "generally \*320 accepted by other courts" to support their decisions, rather than engaging the evidence and subjecting it to a rational, science-based analysis.<sup>130</sup>

In fact, the reliance on "long history of use" seems to reflect a common bias that affects intuitive decision making--namely, the concept of "belief perseverance."<sup>131</sup> This form of bias is "the tendency to maintain existing beliefs in the face of evidence that ought to weaken or even totally reverse those beliefs."<sup>132</sup> According to social psychologists, whatever is learned first seems to have a "primacy effect"--information presented earlier has more influence on judgments than information presented later.<sup>133</sup> Individuals exposed to subsequent, possibly contradictory or conflicting information, disregard the later information, assume it is less reliable or valid, or interpret later evidence in ways that is consistent with their initial impression (perhaps explaining the long history of leech craft).<sup>134</sup> Moreover, the problem of "confirmation bias" causes people to seek out information that supports the original belief and to avoid information that contradicts those beliefs.<sup>135</sup>

The combination of belief perseverance and confirmation bias might explain judges' reluctance to find the so-called matches unsupported--even in the face of ample, compelling testimony that there is absolutely no legitimate support for such conclusions. Consider the contemporary physician writing about leeches, who \*321 muses, "[i]t seems hard to believe that the many educated observers over the centuries were completely wrong in their assessment of clinical improvement following bloodletting."<sup>136</sup> This comment seems to reinforce the powerful effect of belief perseverance, even in the face of abundant data to the contrary.<sup>137</sup>

## VI. The Story Post-NAS Report: One Year Later

So now that the report, with its excoriation of the current state of individualization specialties, has been published for more than a year, it is interesting to see how courts have responded. As of the beginning of May 2010, there are about a dozen cases mentioning or discussing the report--one U.S. Supreme Court case discussing the Confrontation Clause,<sup>138</sup> several federal district court cases, and three state cases--a few of which will be discussed. None of the challenges seeking to exclude forensic science evidence on reliability grounds has succeeded in the court. Two decisions have placed some constraints upon the testimony--allowing the conclusion of a match while limiting or excluding the degree of certainty testimony,<sup>139</sup> and two other federal cases appeared unmoved by the report in analyzing a question of fingerprint comparison admissibility.<sup>140</sup>

In deciding that lab results constituted the "testimonial" statements subject to Confrontation Clause mandates, the Supreme Court in *Melendez-Dias v. Massachusetts* mentioned the report, stating:

"[T]here is wide variability across forensic science disciplines with regard to techniques, methodologies, reliability, types and numbers of potential errors, research, general acceptability, and published material." National Academy Report . . . (discussing problems of subjectivity, bias, and unreliability of common forensic tests such as latent fingerprint analysis, pattern/impression analysis, and toolmark and firearms analysis).

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

Contrary to respondent's and the dissent's suggestion, there is little reason to believe that confrontation will be useless in testing \*322 analysts' honesty, proficiency, and methodology-- the features that are commonly the focus in the cross-examination of experts.<sup>141</sup>

It seems to be a good signal that the U.S. Supreme Court is noticing that "common forensic tests" may be unreliable. In the inferior federal courts, the picture may not be so clear.

In *United States v. Montalvo-Rangel*, the defendant challenged the latent-print examiner's testimony on the basis of the NAS Report. In a one-paragraph discussion, the court overruled the objection with no analysis, only noting that the opinion of the examiner matching the print was given with "great confidence."<sup>142</sup>

In *United States v. Rose*, the trial court found persuasive the "generally accepted by other courts" rationale: "Before and after *Crisp*, it appears that every federal circuit . . . has found expert fingerprint identification testimony admissible . . ." <sup>143</sup> *Rose*, however, does discuss the NAS Report but decides that it did not "conclude that fingerprint evidence was unreliable such as to render it inadmissible . . ." <sup>144</sup> In fact, the judge in *Rose* goes so far as to say that "Judge Harry Edwards, who co-chaired the project, made it clear that nothing in the report was intended to answer the 'question whether forensic science evidence in a particular case is admissible under applicable law.'" <sup>145</sup>

As a result of this type of analysis (and several briefs citing Judge Edwards), Judge Edwards responded pointedly:

If courts blindly follow precedent that rests on unfounded scientific premises, this will lead to unjust results. Nothing in established law compels this course. So when the report was released and I said that judges must continue to follow the law, I did not mean to suggest that judges would apply existing law without taking into account the findings in the report that raise serious doubts about the validity and reliability of certain forensic disciplines and practices.

. . .

The point here is simple: When scientific methodologies once considered sacrosanct are modified or discredited, the judicial system must accommodate the changed scientific landscape.<sup>146</sup>

Two cases involving firearm toolmark comparison engage in a more thoughtful analysis of the meaning of the report, but neither appears to grasp the implication of why a conclusion of a "match" is not currently supportable.

\*323 *United States v. Taylor*, a district court decision from New Mexico, provides a detailed explanation of the shortcomings of firearm toolmark comparison, discussing the lack of standards by which a match is declared, the subjective basis for the conclusion of a match, and the lack of standards for even distinguishing between class, subclass, and individual characteristics.<sup>147</sup> "[T]he . . . theory is circular. An examiner may make an identification when there is sufficient agreement, and sufficient agreement is defined as enough agreement for an identification."<sup>148</sup> The court also quotes the damning language from the NAS Report that, even with better training and new techniques, "the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates."<sup>149</sup>

Nonetheless, the significance of those failures<sup>150</sup> is swept aside as orthogonal to the underlying validity. The court, citing pre-NAS Report cases, finds the practice of cartridge comparison sufficiently reliable to be admitted and permits the expert to give his opinion "within a reasonable degree of certainty in the firearms examination field."<sup>151</sup>

United States v. Willock<sup>152</sup> is a deeper, more thoughtful analysis. The court cites the conclusion expressed in the NRC Ballistic Imaging Report: “The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated,” and “additional general research on the uniqueness and reproducibility of firearms-related toolmarks would have to be done if the basic premises of firearms identification are to be put on a more solid scientific footing.”<sup>153</sup> Yet, in its analysis, the court finds the pre-NAS Report decisions compelling, noting that “the furthest” any court to date has gone is to exclude testimony where examiners’ results were not confirmed or documented; or “to restrict the degree of certainty to which the examiners could express their identifications.”<sup>154</sup> The courts seems to note-- without disapproval-- \*324 that one 2007 federal case allowed the expert to state a conclusion with “100% degree of certainty.”<sup>155</sup>

Quoting Kumho Tire for the principle that the Daubert factors are not “holy writ,” the Willock court does not address the question of whether toolmark identification is science (implicitly failing to recognize that claims of a match are based upon scientific and statistical principles) and largely follows suit with the pre-NAS Report cases.<sup>156</sup> However, Willock goes beyond most courts in limiting the testimony: The trial court requires the prosecution to present testimony only from (a) a qualified examiner; (b) who followed the standard theory (“despite its subjectivity”); and (c) who documented in detail his procedures so as to allow another examiner to follow the original steps.<sup>157</sup> However, the court goes one step further than the pre-NAS Report toolmark cases and restricts the examiner to stating opinions and conclusions “without any characterization as to the degree of certainty with which he holds them.”<sup>158</sup> This case is the first to so limit the testimony and may usher in a generation of new decisions.

While both of the courts recognize that toolmark comparison “does not have sufficient rigor to be evaluated as science,”<sup>159</sup> they categorize this testimony as either technological or other specialized testimony<sup>160</sup> and then proceed to admit the testimony with some limitations. What the courts continue to miss, however, is that the declaration of a match (a conclusion of individualization), is, by its very nature, a scientific finding:

[A] conclusion of individualization implies that the evidence originated from that source, to the exclusion of all other possible sources. The determination of uniqueness requires measurements of object attributes, data collected on the population frequency of variation in these attributes, testing of attribute independence, and calculations of the probability that different objects share a common set of observable attributes.<sup>161</sup>

The judiciary to date still does not seem to understand fully the nature of the problem; while some courts explain the problem quite well, they are unprepared to foreclose the expert from declaring a match. Other courts, like United States v. Rose, have simply proceeded along as if the report was meaningless; perhaps \*325 proving the triumph of belief perseverance in the face of contradictory (and compelling) information.

I suggest that we should encourage judges to work with the language of the NAS Report to write an opinion in which they address the following factors: (1) measurement of object attributes; (2) data on population frequency of variation in the attributes; (3) evidence of attribute independence; and (4) calculation of the probability that different objects share a common set of attributes.

By urging judges to use these factors, which are complex and intricate (and follow the template of DNA comparison), judges will be less inclined to resort to heuristic-based intuitive decision making and will instead rely on deliberative process. In working through the deliberative process, it becomes clear very quickly that these elements cannot be met. Thus, the only response is to exclude the evidence or disallow conclusions of a match. Most likely, the latter is the course courts would take, and in my opinion the correct approach. As is true in both love and science, there is nothing simple about a match.

## VII. Conclusion

With other scholars, I have previously argued that because judges apparently were unwilling to exclude prosecutorial forensic evidence, perhaps judges could follow the “middle way” by limiting, if not excluding, the testimony.<sup>162</sup> My suggestion (and that of others) is primarily to let the expert testify about points of comparison, without giving a conclusion to the jury. This



WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

approach does not resolve the problems identified in the report; but it possibly cures the worst problems with individualization evidence.

One might have hoped that the courts would, at least, be willing to take this step after the NAS Report waved red flags about the problems of forensic individualization and the worrisome implications of those shortcomings. However, it seems apparent in the few cases that have been decided to date that courts are operating predictably with belief perseverance and are simply assimilating the implications of the NAS Report by interpreting the report to conform with their prior beliefs.

Again, however, social science may provide some clues as to how to affect this particular form of cognitive bias so that judges really understand the dangers of admitting conclusions about matches: namely, by requiring greater accountability of judges in their decision making. If subjects are told ahead of time that they will be accountable for their judgments, they are much less susceptible to primacy or belief perseverance.<sup>163</sup> Here the role of the Supreme Court is critical: if the Court continues to recognize the problems in forensic science, as Justices \*326 appeared to in the language of Melendez-Diaz, the inferior federal courts likely will realize that they will be held accountable for their decisions on forensic science and will begin to evaluate the testimony in a more critical, thoughtful fashion.

Moreover, the recent comments of Judge Edwards about the report provide a crucial first step to correcting the course that some courts have chosen. The importance of judges holding the prosecution to legitimate reliability standards cannot be underestimated. To paraphrase Judge Gertner, when life or liberty hangs in the balance, we should expect better forensic science evidence than has been historically demanded. “We should require more,”<sup>164</sup> both from our forensic science experts and from our judiciary.

Footnotes

<sup>a1</sup> © 2010 Jane Campbell Moriarty, Professor of Law & Director of Research and Development, University of Akron School of Law. The author wishes to thank Dan Medwed, Joelle Moreno, Jay Koehler, Frank Yates, Elisa Krackow, David Harris, Jeff Rachlinski, and Adina Schwartz, all of whom provided helpful insights into either law or psychology; Matt Powell for help with tracking down out-of-the-ordinary research; Akron Law students Erick Rigby and Marisa Main for terrific research assistance; The University of Akron School of Law for support for this Article; and the editors of the Utah Law Review for their patience and fine work. All errors remain the author’s.

<sup>1</sup> Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592-93 (1993).

<sup>2</sup> National Research Council of the National Academies, Strengthening Forensic Science in the United States: A Path Forward 7 (2009) [hereinafter NAS Report].

<sup>3</sup> Id. at 108-09 (quoting Jennifer L. Mnookin, The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate, 7 Law, Probability & Risk 127, 134 (2008)).

<sup>4</sup> Id. at 109 (quoting Peter J. Neufeld, The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggestions for Reform, 95 Am. J. Pub. Health 107, 109 (2005)).

<sup>5</sup> Id.

<sup>6</sup> Id.

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

7 Id. at 87.

8 Id. at 109 (stating that the courts have been ineffective in addressing the issue of the validity of forensic science).

9 See id. at 1-33 (explaining the shortcomings of forensic science and providing a summary of sought-for changes in the system).

10 Id. at 4.

11 Id.

12 Id.

13 Id.

14 Id. at 5-6.

15 Id. at 6.

16 Id. at 6-7.

17 Id. at 7-8.

18 Id. at 12.

19 See id.

20 Id.

21 Id.

22 Id.

23 See id. at 87 (“The degree of science in a forensic science method may have an important bearing on the reliability of forensic evidence in criminal cases.”). For a searching empirical analysis explaining the relationship between forensic science errors and wrongful convictions, see Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 Va. L. Rev. 1 (2009).

24 The foundations of arson investigation, erroneously believed by many to rest on firm scientific footing, suffers from multiple shortcomings. The NAS Report discusses the “paucity of research” and states that “many of the rules of thumb that are typically assumed to indicate an accelerant was used ... have shown not to be true.” NAS Report, *supra* note 2, at 173.

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

- 25 Id. at 136-64. See generally Jane Campbell Moriarty and Michael Saks, Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping, 44 Judges' J. 16, 17 (2005) (explaining that "individualization specialties seek to associate an item of evidence found at a crime scene with its unique source, to the exclusion of all others"). For further discussion of the concept of "individualization," see John I. Thornton & Joseph L. Peterson, The General Assumptions and Rationale of Forensic Identification, in Modern Scientific Evidence: The Law And Science of Expert Testimony §§ 31:9-31:11 (David L. Faigman et al. eds., 2005).
- 26 Id.
- 27 See id. at 41 (stating DNA analysis, "with its well-defined precision and accuracy, has set the bar higher for other forensic science methodologies").
- 28 See id. at 43-44.
- 29 Id. at 87.
- 30 Id. at 44.
- 31 See Part VI, *infra*, discussing post-NAS Report cases. Even since the report was issued, many federal prosecutors have continued to argue the evidence is completely reliable and the NAS Report should not be taken into consideration. See Harry T. Edwards, The National Academy of Sciences Report on Forensic Sciences: What it Means for the Bench and Bar, presentation at the Superior Court of the District of Columbia, May 6, 2010, at 4.
- 32 NAS Report, *supra* note 2, at 42.
- 33 See Pierre Schlag, Commentary Law and Phrenology, 110 Harv. L. Rev. 877, 880 (1997) (discussing this methodology as used in phrenology). In *Misconvictions, Science and the Ministers of Justice*, I discuss the methodological parallels between phrenology and individualization specialties. 86 Neb. L. Rev. 1, 17-19 (2007).
- 34 It is important to distinguish legal precedent from this "generally accepted by other courts" principle. The former refers to the prior court rules of law; the latter refers to findings of scientific reliability of evidence.
- 35 Jennifer L. Mnookin, The Validity of Latent Fingerprint Identification: Confessions of a Fingerprint Moderate, 7 Law, Probability & Risk 127, 134 (2008) ("We do have a roughly 100-year long, extremely informal 'natural experiment' as a result of [fingerprinting's] quite substantial investigative use. Whenever a fingerprint examiner matches a latent print to a suspect, and then, subsequently, independent evidence emerges to tie the suspect to the location where the print was left, this new evidence corroborates the correctness of the fingerprint identification."). Although this may be true, we are still left with the nagging lack of certainty about the meaning and importance of a match; whether in this case the match was correct; and equally significant, how often is the claim of a match not correct? As Professor Mnookin notes, in a 1995 proficiency test, nearly 20 percent of the test takers found a match where none existed. Id. at 135. For more on error rates among fingerprint examiners, see Simon A. Cole, Grandfathering Evidence: Fingerprint Admissibility Rules from Jennings to Llera Plaza and Back Again, 41 Am. Crim. L. Rev. 1189 (2004) and Simon A. Cole, More than Zero: Accounting for Error in Latent Print Identifications, 95 J. Crim. L. & Criminology 985 (2005).
- 36 See Robert Burton, *On Being Certain: Believing You Are Right Even When You're Not* (2008) (explaining that the certainty of a hunch may well be a misplaced concept). Burton explains, "[i]nternal bias and a misplaced feeling of knowing routinely overpower and outsmart the intellect." Id. at 149. However, we have "no mechanism for establishing the accuracy of a line of reasoning until

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

it has produced a testable idea.” Id. at 151. Judges might well think about this concept when they continue to “feel” as though fingerprints, handwriting, and toolmarks can be accurately “matched.”

- <sup>37</sup> Confirmation Hearing on the Nomination of John G. Roberts, Jr. to be Chief Justice of the United States: Hearing Before the S. Comm. on the Judiciary, 109th Cong. 55 (2005) (statement of John G. Roberts, Jr., nominee to be Chief Justice of the United States); see also Jeffrey Toobin, No More Mr. Nice Guy: The Supreme Court’s Stealth Hard-liner, *The New Yorker*, May 25, 2009, at 42. Professor Toobin, however, disagrees with Justice Roberts’s self-assessment, stating that even more than Justice Scalia, “Roberts has served the interests, and reflected the values, of the contemporary Republican Party.” Id. at 44.
- <sup>38</sup> Frederick Schauer, Do Cases Make Bad Law?, 73 *U. Chi. L. Rev.* 883, 887-88 (2006). It is far beyond the scope of this Article to review the long history of jurisprudence concerning the judiciary as lawmakers. But as Professor Schauer quite accurately comments, it is “far too late in the day to deny that judges are often (some would say ‘always’) engaged in the process of making law.” Id. at 888 (citing Duncan Kennedy, *Legal Formality*, 2 *J. Legal Stud.* 351, 378 (1973)).
- <sup>39</sup> Richard A. Posner, The Role of the Judge in the Twenty-First Century, 86 *B.U. L. Rev.* 1049, 1051 (2006). Many other scholars have roundly disagreed with the now-Chief Justice’s view of the role of the judiciary, for various and often competing reasons. See Michael P. Allen, A Limited Defense of (at Least Some of) the Umpire Analogy, 32 *Seattle U. L. Rev.* 525, 526 n.4 (2009).
- <sup>40</sup> See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999); *Gen. Elec., Co. v. Joiner*, 522 U.S. 136 (1997); *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993).
- <sup>41</sup> See, e.g., *Fed. R. Evid.* 702 (providing, in pertinent part, that an expert may testify if “(1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case”); accord *Kumho Tire Co.*, 526 U.S. at 141; *Gen. Elec. Co.*, 522 U.S. at 146-47; *Daubert*, 509 U.S. at 589-90 (requiring sufficient proof that the proposed expert evidence is sufficiently reliable to be admitted). Although Federal Rule of Evidence 702 was amended to add reliability requirements, there is much agreement that the rule reflects the development in the common law. Federal Rule of Evidence 702 was “amended in response to *Daubert* ... and to the many cases applying *Daubert*, including *Kumho Tire*.” *Fed. R. Evid.* 702 advisory committee’s note on 2000 Amendments.
- <sup>42</sup> The “Frye general acceptance test” requires proof that the novel scientific evidence is generally accepted in the field to which it belongs. The general acceptance test does not analyze the reliability of the proposed evidence; it asks whether novel scientific evidence has reached the tipping point at which it has become generally accepted by scientists in the field. See *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923). Many federal cases substantially incorporate some variant of the Frye test in their reliability determination, because it is one of the factors listed by the *Daubert* court. See 509 U.S. at 588-89, 598. For a state-by-state breakdown of which states use the respective tests, see 2 Jane Campbell Moriarty, *Psychological and Scientific Evidence in Criminal Trials* app. 1A (2008) (providing a state-by-state analysis of admissibility standards) [hereinafter Moriarty, *Criminal Trials*].
- <sup>43</sup> D. Michael Risinger, Defining the “Task At Hand”: Non-Science Forensic Science After *Kumho Tire Co. v. Carmichael*, 57 *Wash. & Lee L. Rev.* 767, 773 (2000) (citing *Kumho Tire Co.*, 526 U.S. at 141). This point is important, again, for distinguishing between precedent and decisions about reliability of evidence.
- <sup>44</sup> See *Kumho Tire Co.*, 526 U.S. at 150 (for the principle that the inquiry is “a flexible one”) (citing *Daubert*, 509 U.S. at 594).
- <sup>45</sup> See *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1316 (9th Cir. 1995) (On remand, Judge Kozinski, writing for the court, applied the standard created by the Supreme Court with the statement, “[m]indful of our position in the hierarchy of the federal judiciary, we take a deep breath and proceed with this heady task.”). Many states, through their supreme courts’ decisions, also have either adopted or rejected *Daubert*, in whole or in part. See Moriarty, *Criminal Trials*, supra note 42, at app. 1A. Thus, very



WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

clearly, courts both federal and state are engaged in rule-making in the area of expert testimony.

<sup>46</sup> Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 589 n.7 (1993) (mentioning that Fed. R. Evid. 702 is the source of the judge's gatekeeping duty).

<sup>47</sup> During the ten years I worked on my treatise, Moriarty, Criminal Trials, supra note 42, my impression was that courts were not requiring the prosecution to comply with the trilogy requirements. Of course, that was just my impression. Better evidence comes from the studies cited infra at note 48.

<sup>48</sup> For empirical data establishing that judges permit much prosecutorial expert testimony and simultaneously disallow much of defense expert testimony, see D. Michael Risinger, Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?, 64 Alb. L. Rev. 99 (2000). Recently, Professor Risinger analyzed all reported decisions on defense challenges to fingerprint evidence and concluded that, overwhelmingly, the courts rejected these challenges--not on the basis of an accurate reliability analysis but for a variety of other, less data-driven reasons, noting that "there is some reason to believe that judges as a group are resistant to rejecting prosecution proffers of expert testimony." 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, 473 (2007).

<sup>49</sup> NAS Report, supra note 2, at 11 (emphasis in original). Of course, the universe of reported decisions is far smaller than that of all decisions courts make, a point the NAS Report concedes. *Id.* Nonetheless, if the reported cases are at all representative, the discrepancy is noteworthy.

<sup>50</sup> This point has been stated elsewhere: "There is almost no [prosecutorial] expert testimony so threadbare that it will not be admitted if it comes to a criminal proceeding under the banner of forensic science." Moriarty & Saks, supra note 25, at 29.

<sup>51</sup> See Deirdre Dwyer, (Why) are Civil and Criminal Expert Evidence Different?, 43 Tulsa L. Rev. 381, 384-87 (2007).

<sup>52</sup> *Id.* at 384-85.

<sup>53</sup> *Id.* at 383. Dwyer does acknowledge the limitations of the studies, pointing out that "[t]here are significant methodological difficulties with inferring general trial conduct from reported decisions." *Id.*

<sup>54</sup> See *id.* at 385 (noting that "[t]here are generally greater consequences, socially and physically, of criminal conviction compared with adverse civil judgment, and so we might expect that the rules of evidence in a criminal action are geared more towards reducing the risk of an erroneous outcome than they are in a civil action").

<sup>55</sup> The following cases provide a small sample of courts' in-depth opinions on toxic tort expert evidence: *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 143-47 (1997) (analyzing admissibility of plaintiffs' proposed expert testimony that exposure to PCBs promoted decedent's lung cancer); *McLain v. Metabolife Int'l, Inc.*, 401 F.3d 1233, 1239-52 (11th Cir. 2005) (reversing jury verdict for plaintiffs because plaintiffs' expert evidence that ingestion of defendant product for weight loss was a likely cause of their respective strokes and heart attack was not reliable); *Norris v. Baxter Healthcare Corp.*, 397 F.3d 878, 883-87 (10th Cir. 2005) (explaining why plaintiffs' experts' opinions that silicone breast implants caused autoimmune diseases was not reliably grounded on existing data); *Moore v. Ashland Chem. Inc.*, 151 F.3d 269, 274-79 (5th Cir. 1998) (upholding trial court's decision to exclude expert evidence on allegations that plaintiff's exposure to toluene caused his respiratory disease); *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1316-22 (9th Cir. 1995) (on remand from the U.S. Supreme Court, analyzing whether plaintiffs' claim that Bendectin exposure to fetuses in utero caused their limb malformation); *Henricksen v. Conoco Phillips Co.*, 605 F. Supp. 2d 1142, 1152-79 (E.D. Wash. 2009) (analyzing expert testimony that plaintiff's exposure to benzene caused the development of his

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

myelogenous leukemia); *Amorgianos v. Nat'l R.R. Passenger Corp.*, 137 F. Supp. 2d 147, 160-191, (E.D.N.Y. 2001), *aff'd*, 303 F.3d 256 (2d Cir. 2002) (explaining why experts' opinions that exposure to xylene vapor glue fumes caused central nervous system and peripheral nervous system injury was not sufficiently reliable to be admissible).

<sup>56</sup> See cases cited *supra* note 55. In *Uncertainty and Informed Choice: Unmasking Daubert*, 104 Mich. L. Rev. 257, 258-67 (2005), Professors Margaret A. Berger and Aaron D. Twersky argue that toxic tort plaintiffs have a great deal of difficulty, post-Daubert, in meeting the courts' requirements concerning expert testimony.

<sup>57</sup> NAS Report, *supra* note 2, at 96.

<sup>58</sup> *Id.* at 106.

<sup>59</sup> See cases mentioned *infra* at notes 60-62.

<sup>60</sup> See, e.g., *United States v. Mitchell*, 365 F.3d 215, 245-46 (3d Cir. 2004) (allowing presentation of fingerprint comparison but allowing the defense to call "counter-experts" to raise doubt about the analysis); *United States v. Crisp*, 324 F.3d 261, 269-70 (4th Cir. 2003) (basing its decision that fingerprint evidence is reliable, at least in part, on the previous decades of consistent admission of such evidence); *United States v. Havvard*, 260 F.3d 597, 600 (7th Cir. 2001) (holding that fingerprint comparison met Daubert's reliability standard); *United States v. Rogers*, 26 F. App'x 171, 173 (4th Cir. 2001) (allowing fingerprint comparison evidence while seemingly shifting the burden on the defendant to show that the analysis is unreliable); *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 575 (E.D. Pa. 2002) (holding that because fingerprint evidence is "sufficiently reliable for an English court," it must also be reliable under Rule 702). For an excellent analysis of the shortcomings of fingerprint comparison based upon the twin pillars of uniqueness and individualization, see Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 *Law, Probability & Risk* 233, 242-46 (2009).

<sup>61</sup> See, e.g., *United States v. Glynn*, 578 F. Supp. 2d 567, 574-75 (S.D.N.Y. 2008) (acknowledging that firearm toolmark comparison "lacks the rigor of science" and "suffers from greater uncertainty than many other kinds of forensic evidence," yet allows testimony that a firearms match was "more likely than not"); *United States v. Greene*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005) (allowing expert to testify about the points of similarity between a shell casing at the scene and one from defendant's weapon but disallowing any testimony that the shell casings matched); *United States v. Hines*, 55 F. Supp. 2d 62, 70-71 (D. Mass. 1999) (limiting testimony of handwriting comparison witness to discussing points of similarity but disallowing conclusion testimony); *United States v. Santillan*, 1999 WL 1201765, 5 (N.D. Cal., Dec. 3, 1999). Post-NAS Report cases are discussed, *infra*, in Part IV.

<sup>62</sup> See, e.g., *United States v. Adeyi*, 164 F. App'x 944, 946 (2d Cir. 2006) (upholding admission of handwriting analysis relying almost exclusively on precedent); *United States v. Crisp*, 324 F.3d 261, 271-72 (4th Cir. 2003) (affirming decision by the trial court that expert testimony regarding handwriting comparisons and conclusion of authorship were reliable under the Daubert standard); *United States v. Battle*, 1999 WL 596966, 4 (10th Cir. Aug. 6, 1999) (upholding the trial court's decision that expert testimony about a conclusion of handwriting comparison met the standard of reliability and relevancy); *United States v. Gonzales*, 90 F.3d 1363, 1371 (8th Cir. 1996) (finding that the district court was within its discretion to admit handwriting comparison evidence). For a full explanation of the history of legal challenges concerning the foundation of handwriting comparison, see *Risinger, Goodbye to All That*, *supra* note 48, at 457-67.

<sup>63</sup> Innocence Project, *Wrongful Convictions Involving Unvalidated or Improper Forensic Science that Were Later Overturned Through DNA Testing* (2009), [http://www.innocenceproject.org/docs/DNA\\_Exonerations\\_Forensic\\_Science.pdf?phpMyAdmin=52c4ab7ea46t7da4197](http://www.innocenceproject.org/docs/DNA_Exonerations_Forensic_Science.pdf?phpMyAdmin=52c4ab7ea46t7da4197); Innocence Project, *Factors Leading to Wrongful Convictions*, [http://www.innocenceproject.org/understand/factors\\_74\\_chart.php](http://www.innocenceproject.org/understand/factors_74_chart.php) (last visited June 1, 2010).

<sup>64</sup> See, e.g., *United States v. Santiago*, 156 F. Supp. 2d 145, 151-52 (D.P.R. 2001) (admitting hair comparison analysis after

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

referencing persuasive authority and independent research); *Williamson v. Ward*, 110 F.3d 1508, 1522-23 (10th Cir. 1997) (affirming on other grounds but rejecting the district court's ruling that hair analysis was not admissible); *U.S. v. Matta-Ballesteros*, 71 F.3d 754, 766-67 (9th Cir. 1995) (approving of the admission at trial of expert testimony on hair comparison). For critical commentary on microscopic hair comparison, see the NAS Report, *supra* note 2, at 155-61, and Edwards, *supra* note 31, at 6 (stating that "hair comparisons without mitochondrial DNA are highly questionable").

65 NAS Report, *supra* note 2, at 8 ("The simple reality is that the interpretation of forensic evidence is not always based on scientific studies to determine its validity. This is a serious problem. Although research has been done in some disciplines, there is a notable dearth of peer-reviewed, published studies establishing the scientific bases and validity of many forensic methods.").

66 *United States v. Havvard*, 117 F. Supp. 2d 848, 854 (S.D. Ind. 2000), *aff'd*, 260 F.3d 597 (7th Cir. 2001).

67 *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 575-76 (E.D. Pa. 2002), *vacating*, 179 F. Supp. 2d 492 (E.D. Pa. 2002).

68 *Id.* at 576.

69 *Id.*

70 *Id.* at 572.

71 *United States v. Prime*, 220 F. Supp. 2d 1203, 1210 (W.D. Wash. 2002), *vacated and remanded on other grounds*, 543 U.S. 1101 (2005), *subsequently aff'd on this issue*, 431 F.3d 1147 (9th Cir. 2005).

72 *Prime*, 431 F.3d at 1153.

73 *Id.* (noting that in one professor's studies, experts arrived at the correct conclusions 87 percent of the time). As noted by the Llera Plaza court (and echoed by numerous scholars), the proficiency tests are not rigorous: "[T]he proficiency tests are less demanding than they should be. To the extent that this is the case, it would appear that the tests can be of little assistance in providing the test makers with a discriminating measure of the relative competence of the test takers." 188 F. Supp. 2d at 565. For further information on proficiency tests of fingerprint examiners, see NAS Report, *supra* note 2, at 143-44.

74 *People v. Davis*, 710 N.E.2d 1251, 1256 (Ill. App. Ct. 1999), *conviction vacated*, No. 94 CF 76, 2006 WL 2641753 (Cir. Ct. Ill. Mar. 7, 2006).

75 See *id.* at 1258-59.

76 *People v. Davis*, No. 94 CF 76, 2006 WL 2641753 (Cir. Ct. Ill. Mar. 7, 2006) (holding that defendant was denied his constitutional right to effective counsel and overturning the defendant's conviction).

77 See *Davis*, 710 N.E.2d at 1258-59.

78 On the other hand, this case may not be that much of an outlier because federal courts have permitted much expert testimony in criminal cases that appears to have dicey scientific grounding, including profile testimony about specific types of pedophiles. See,

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

e.g., *U.S. v. Hitt*, 473 F.3d 146, 151, 158-59 (5th Cir. 2006) (upholding admission of expert testimony “to explain the behavior of those accused of sexual offenses”). Federal courts also have admitted allegedly scientific comparisons between photographs of items and the items themselves. See, e.g., *U.S. v. McKreith*, 140 F. App’x 112, 116 (11th Cir. 2005) (allowing expert testimony on the comparison of the defendant’s shirt that was seized in his apartment with the shirts depicted in bank surveillance images).

79 See Erica Beecher-Monas, *Reality Bites: The Illusion of Science in Bite-Mark Evidence*, 30 *Cardozo L. Rev.* 1369, 1375-88 (2009); NAS Report, *supra* note 2, at 173-76.

80 See Beecher-Monas, *supra* note 79, at 1378-84.

81 See *id.* at 1385-87.

82 See, e.g., *State v. Garrison*, 585 P.2d 563, 566 (Ariz. 1978) (affirming trial court’s decision admitting expert testimony that the probability of two sets of teeth being identical was eight in one million, based on “articles written throughout the literature that do mention the possibility or the numerical values of finding two (sets of teeth) that match.”).

83 Beecher-Monas, *supra* note 79, at 1386.

84 See A.P.A. Broeders, *Of Earprints, Fingerprints, Scent Dogs, Cot Deaths, and Cognitive Contamination--A Brief Look at the Present State of Play in the Forensic Arena*, 159 *Forensic Sci. Int’l* 148, 153-54 (2006) (explaining why forensic science individualization does not meet the scientific standard).

85 See NAS Report, *supra* note 2, at 87 (“[N]o forensic method other than nuclear DNA analysis has been rigorously shown to have the capacity to consistently and with a high degree of certainty support conclusions about ‘individualization’ (more commonly known as ‘matching’ of an unknown item of evidence to a specific unknown source).”).

86 See NAS Report, *supra* note 2, at 21 (noting the need for standardized testimony).

87 See Dawn McQuiston-Surrett & Michael J. Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 *Hastings L.J.* 1159, 1188 (2008) (explaining studies conducted to determine how jurors analyze such terms as “match” and “consistent with” and finding that jurors overestimated the meaning of such terms); accord NAS Report, *supra* note 2, at 21 (“[U]se of such terms can and does have a profound effect on how the trier of fact ... perceives and evaluates scientific evidence.”).

88 See NAS Report, *supra* note 2, at 176 (“[T]here is considerable dispute about the value and reliability of the collected data for interpretation.”).

89 See cases collected in Beecher-Monas, *supra* note 79, at 1372-74, and accompanying notes.

90 See *United States v. Crisp*, 324 F.3d 261, 271-72 (4th Cir. 2003).

91 *Id.* at 268-69.



WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

- 92 Id. at 269.
- 93 Id. at 271.
- 94 See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 152 (1999) (“[A] trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony.”).
- 95 *Crisp*, 324 F.3d at 273 (Michael, J., dissenting).
- 96 NAS Report, *supra* note 2, at 139 (noting that the threshold for making a source identification is “deliberately kept subjective”).
- 97 Id.
- 98 Id. at 143 n.31 and accompanying text.
- 99 Id. at 143-44.
- 100 Id. at 166.
- 101 NAS Report at 161.
- 102 Id. at 96-97.
- 103 Id. at 110.
- 104 Id.
- 105 *Mnookin*, *supra* note 35, at 141.
- 106 See cases cited *supra* note 55. Granted, part of the courts’ apparent comfort in the civil cases may arise from the parties’ well-structured challenges and responses to the expert testimony. Many criminal practitioners are not developing the challenges as well as their civil counterparts. Nonetheless, serious, science-based challenges have been raised against fingerprints, toolmarks, and handwriting comparison in many criminal cases, and nearly universally courts have rejected them. See cases cited *supra* notes 60-64.
- 107 *Daubert*, 509 U.S. 579, 590 n.9 (describing the standard as “evidentiary reliability--that is, trustworthiness”).
- 108 See T.S. Eliot, *Little Gidding*, in *The Complete Poems and Plays* 142 (Harcourt Brace & World, Inc., 1971).
- 109 I.S. Whitaker et al., *Hirudo Medicinalis: Ancient Origins of, and Trends in the Use of Medicinal Leeches Throughout History*, 42 *British J. of Oral & Maxillofacial Surgery* 133, 134-36 (2004); accord N. Papavramidou & H. Christopoulou-Aletra, *Medicinal Use*

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

of Leeches in the Texts of Ancient Greek, Roman and Early Byzantine Writers, 39 *Internal Medicine J.* 624, 624-27 (2009) (discussing the historical origins of the medicinal use of leeches). Others claim bloodletting as a form of treatment has a 3,000-year history. See Gilbert R. Seigworth, *Bloodletting over the Centuries*, 80 *N.Y. St. J. Med.* 2022-28 (Dec. 1980) (citing the use of bloodletting by the ancient Egyptians). In any event, the use of leech bleeding has a long history indeed. David Faigman also has mentioned the leech-bleeding analogy. David L. Faigman, *Anecdotal Forensics, Phrenology, and Other Abject Lessons from the History of Science*, 59 *Hastings L.J.* 979, 985 (2008).

110 Seigworth, *supra* note 109, at 2024.

111 Recently, scientists have discovered the important anticoagulant properties of leeches and have successfully used leeches in microsurgical application. Whitaker, *supra* note 109, at 136. However, the use of leeches for simply “bleeding cures” for various disease processes turns out not to have much foundation.

112 See Richard Bernard, *Guide to Grand Jury Men* (1627); William Perkins, *Discourse on the Damned Art of Witchcraft* (1608); see also Frances Hill, *The Salem Witch Trials Reader* 3 (2000).

113 Jane Campbell Moriarty, *Wonders of the Invisible World: Prosecutorial Syndrome and Profile Testimony in the Salem Witchcraft Trials*, 26 *Vt. L. Rev.* 43, 57-63 (2001) (discussing the expert authority on which judges relied to detect witchcraft).

114 Hugh Bowden, *Classical Athens and the Delphic Oracle, Divination and Democracy* 19 (2005).

115 *Id.* at 29-30.

116 *Id.* at 26-28. It cannot be lost on the reader that many superstitious beliefs persist despite the absence of accuracy. Today, some remnants of the bloodletting years continue to hold sway--recently the use of “cupping”--a form of “dry bloodletting”--has found favor with some alternative medicine fans. For more on its historical use, see Seigworth, *supra* note at 109. For a fuller explanation of modern-day cupping, see *Welcome to the British Cupping Society*, <http://www.britishcuppingsociety.org/Portal/index.php> (last visited June 1, 2010).

117 Bowden, *supra* note 114, at 19.

118 See Richard A. Posner, *How Judges Think* (2008); Cass Sunstein, et al., *Are Judges Political? An Empirical Analysis of the Federal Judiciary* (2006); *The Psychology of Judicial Decision Making* (David Klein and Gregory Mitchell eds., 2010); Chris Guthrie, Jeffrey J. Rachlinski & Andrew Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 *Cornell L. Rev.* 1 (2007) [hereinafter *Blinking on the Bench*]; Chris Guthrie, Jeffrey J. Rachlinski & Andrew Wistrich, *Inside the Judicial Mind*, 86 *Cornell L. Rev.* 777 (2001); Chad M. Oldfather, *Writing, Cognition, and the Nature of the Judicial Function*, 96 *Geo. L.J.* 1283 (2008); Frederick Schauer, *Do Cases Make Bad Law?*, 73 *U. Chi. L. Rev.* 893 (2006); Dan Simon, *A Psychological Model of Judicial Decision Making*, 30 *Rutgers L.J.* 1 (1998).

119 For more on this idea, see Chad M. Oldfather, *Judges and Humans: Interdisciplinary Research and the Problems of Institutional Design*, 36 *Hofstra L. Rev.* 125, 128 n.11 (2007) (collecting articles referencing the different ways that judges are subject to the same frailties as other humans). Jeff Rachlinski also reminded me that many judges have experience as prosecutors: this background may lead them to think of prosecutorial experts as “tools to build [the] cases” rather than as partisans. (e-mail on file with author).

120 *Blinking on the Bench*, *supra* note 118, at 6-9. For a more detailed discussion of this distinction, see D. Kahneman, P. Slovic & A.

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

Tversky, Judgment Under Uncertainty: Heuristics and Biases (1982); T. Gilovich, D. Griffin & D. Kahneman, Heuristics and Biases: The Psychology of Intuitive Judgment (2002). For a more precise explanation of the distinction, see Steven A. Sloman, Two Systems of Reasoning, in Gilovich et al., supra note 120, at 379. Other scholars have written about the dual process model of decision making. In the area of scientific evidence law, Joseph Sanders has written a most compelling and instructive article. See Joseph Sanders, Kumho and How We Know, 64 L. & Contemp. Probs. 373, 393 (2001) (describing the concepts of “experiential processing” and “rational processing”).

<sup>121</sup> Blinking on the Bench, supra note 118, at 7.

<sup>122</sup> Id. (quoting Shane Frederick, Cognitive Reflection and Decision Making, 19 J. Econ. Persp. 25, 26 (2005)).

<sup>123</sup> Id. at 27.

<sup>124</sup> Id. at 36.

<sup>125</sup> Id. at 10-13 (discussing the Cognitive Reflection Test Model from Shane Frederick, Cognitive Reflection and Decisionmaking, supra note 122, at 26-28, and the judges’ performance on the test).

<sup>126</sup> Thomas Gilovich and Dale Griffin, Introduction--Heuristics and Biases: Then and Now, in Judgment Under Uncertainty, supra note 120, at 7.

<sup>127</sup> See Benjamin R. Newell, et al., The Psychology of Decision Making, 71 (2007) (citing the oft-quoted Kahneman, D. & Tversky, A., The Simulation Heuristic, in Kahneman et al., supra note 120; see also Burton, supra note 36, at 149).

<sup>128</sup> This does not mean that error cannot occur in the deliberative process or that judges always get the right answer. All it means, according to research psychology, is that the probability of error is reduced using this method.

<sup>129</sup> Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).

<sup>130</sup> United States v. Crisp, 324 F.3d 261 (4th Cir. 2003) (majority admitting the evidence uses these intuitive heuristics; the dissent, disallowing the evidence, engages in a long, deliberative analysis).

<sup>131</sup> See Richard Nesbitt & Lee Ross, Human Inference: Strategies and Shortcomings of Social Judgment 18-23, 24-42, 45-53, 462 (1980).

<sup>132</sup> Philip E. Tetlock, Accountability and the Perseverance of First Impressions, 46 Soc. Psychol. Q. 285, 285 (1983).

<sup>133</sup> Id. at 286.

<sup>134</sup> Id. Another bias that might be at work here is the so-called “sunk cost” fallacy, often referred to as “throwing good money after bad.” See Hal R. Arkes & Catherine Blumer, The Psychology of Sunk Cost, 35 Org. Behav. & Hum. Decision Processes, 124-40 (1985). This theory describes an individual’s unwillingness to withdraw from an endeavor after investing money, time, or effort. This bias explains why investors who lose a great deal of money may not be willing to cut their losses and is one explanation for

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

judges' unwillingness to recognize that much of their prior decision making was premised on erroneous beliefs--they are simply "too invested" in their prior decisions to back out. *Id.*

Other articles have discussed the role of cognitive bias in the creation of and decisions about forensic science, including expectation bias, confirmation bias, and so forth. See Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 *Wisc. L. Rev.* 291; Jane Campbell Moriarty, "Misconvictions," *Science*, and the Ministers of Justice, 86 *Neb. L. Rev.* 1, 8, 25 (2007); Michael Risinger et al., *The Daubert/Kumho Tire Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestions*, 90 *Cal. L. Rev.* 1 (2002).

<sup>135</sup> Karl Ask and Pär Anders Granhag, *Motivational Sources of Confirmation Bias in Criminal Investigations: The Need for Cognitive Closure*, 2 *J. Investig. Psychol. Offender Profil.* 43, 45 (2005) (discussing aspects of confirmation bias).

<sup>136</sup> Seigworth, *supra* note 109, at 2027.

<sup>137</sup> In fact, in comparisons of clinical judgments and actuarial judgments (controlled studies), clinical judgments fare more poorly. One scholar notes that "[f]ailure to accept a large and consistent body of scientific evidence over unvalidated personal observation may be described as a normal human failing or, in the case of professionals who identify themselves as scientific, plainly irrational." Robyn M. Dawes et al., *Clinical Versus Actuarial Judgment*, in Kahneman et al., *supra* note 120, at 716, 727.

<sup>138</sup> See *Melendez-Diaz v. Massachusetts*, 129 S. Ct. 2527, 2537 n.6 (2009) (noting the report's conclusion that the forensic science system is badly flawed and refuting the suggestion that forensic examiners are "uniquely reliable").

<sup>139</sup> See *United States v. Willock*, Criminal No. WDQ-08-0086, 2010 U.S. Dist. LEXIS 27473, \*25 (D. Md. Mar. 23, 2010); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1177 (D.N.M. 2009).

<sup>140</sup> See *United States v. Montalvo-Rangel*, No. SA-10-CR-64-XR, 2010 WL 1484708 (W.D. Tex. Apr. 5, 2010); *United States v. Rose*, 672 F. Supp. 2d 723 (D. Md. 2009).

<sup>141</sup> *Melendez-Diaz*, 129 S. Ct. at 2538 (2009) (holding that the Confrontation Clause applies to laboratory reports in criminal cases and discussing the findings of the NAS Report).

<sup>142</sup> *Montalvo-Rangel*, 2010 WL 1484708, at \*3.

<sup>143</sup> *Rose*, 672 F. Supp. 2d at 725.

<sup>144</sup> *Id.*

<sup>145</sup> *Id.*

<sup>146</sup> *Edwards*, *supra* note 31, at 6-7.

<sup>147</sup> 663 F. Supp. 2d 1170, 1177-78 (D.N.M. 2009).

<sup>148</sup> *Id.* at 1177.



<sup>149</sup> Id. at 1178 (citing the NAS Report, *supra* note 2, at 5-20).

<sup>150</sup> These are not the only shortcomings the court notes. While proficiency tests have been conducted, none of them was done as a blind test, which raises doubts about the value of the tests. Id. at 1176.

<sup>151</sup> Id. at 1180.

<sup>152</sup> *United States v. Willock*, Criminal No. WDQ-08-0086, 2010 U.S. Dist. LEXIS 27473, \*25 (D. Md. Mar. 23, 2010).

<sup>153</sup> Id. at \*15 (quoting National Research Council, *Ballistic Imaging 3* (2008)).

<sup>154</sup> Id. at \*16 (citing *United States v. Monteiro*, 407 F. Supp. 2d 351, 374 (D. Mass. 2006); *United States v. Taylor*, No. CR 07-1244, 2009 WL 3347485, at \*9 (D.N.M. Oct. 9, 2009); *United States v. Glynn*, 578 F. Supp. 2d 567, 570 (S.D.N.Y. 2008)). Curiously, the court does not mention here *United States v. Green*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005), in which the court disallowed any conclusions about a match, although it references the opinion elsewhere. See *Willock*, 2009 WL 3617748, at \*14 (discussing *Green* in the context of admissibility concerns of firearm toolmark identification testimony).

<sup>155</sup> See *Willock*, 2009 WL 3617748, at \*16 (citing *United States v. Natson*, 469 F. Supp. 2d 1253, 1261-62 (M.D. Ga. 2007)).

<sup>156</sup> See *id.* at \*19.

<sup>157</sup> Id. at \*20.

<sup>158</sup> Id. at \*24.

<sup>159</sup> *United States v. Taylor*, 663 F. Supp. 2d 1170, 1179 (D. N.M. 2009) (quoting *United States v. Glynn*, 578 F. Supp. 2d 567, 570 (S.D.N.Y. 2008)).

<sup>160</sup> See *United States v. Mouzone*, Criminal No. WDQ-08-086, 2009 WL 3617748, at \*19 (D. Md. Oct. 29, 2009).

<sup>161</sup> NAS Report, *supra* note 2, at 43-44.

<sup>162</sup> Moriarty, *supra* note 134, at 39-41; Moriarty & Saks, *supra* note 25, at 29. Additional suggestions include greater appointment of defense experts and more testimony about actual error rates. For a fuller discussion, see Moriarty *supra* note 134, at 40-41.

<sup>163</sup> Tetlock, *supra* note 132, at 290-91 (suggesting that pre-exposure accountability information may reduce the primacy effect).

<sup>164</sup> *United States v. Green*, 405 F. Supp. 2d 104, 109 (D. Mass. 2005).

WILL HISTORY BE SERVITUDE?: THE NAS REPORT ON..., 2010 Utah L. Rev. 299

---

2010 UTLR 299

---

End of Document

© 2018 Thomson Reuters. No claim to original U.S. Government Works.

# { SEQ CHAPTER \h \r 1}FORDHAM

## University School of Law

Lincoln Center, 150 West 62nd Street, New York, NY 10023-7485

Daniel J. Capra  
Philip Reed Professor of Law

Phone: (b) (6)  
e-mail: (b) (6)

Memorandum To: Rule 702 Subcommittee  
From: Daniel J. Capra, Reporter  
Re: Forensic Evidence, *Daubert* and Rule 702  
Date: June 4, 2018

Here are some thoughts about possible rule amendments to address the two issues that the Subcommittee is considering:

### *1. Amendment to address overstatement by experts:*

#### **Rule 702. Testimony by Expert Witnesses**

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

- (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data;
- (c) the testimony is the product of reliable principles and methods; ~~and~~
- (d) the expert has reliably applied the principles and methods to the facts of the case; and
- (e) the witness does not overstate the probative value of any opinion.

***2. An amendment to address overstatement by feature-comparison experts only:***

**Rule 707. Testimony by Expert Witnesses on Feature-Comparison.**

If a witness qualified as an expert is testifying on the basis of an examination conducted to determine whether an evidentiary sample is similar or identical to a source sample, the expert must:

a) satisfy all the requirements of Rule 702; and

b) accurately state, on the basis of adequate empirical evidence, the meaning of any similarity or match between the evidentiary sample and the source sample, including the rate of error when the witness's method is applied.

***3. An amendment to emphasize that adequacy of basis and reliability of application are matters of admissibility and not weight.***

**Rule 702. Testimony by Expert Witnesses.**

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise, if the court finds by a preponderance of the evidence that the following requirements are met:

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

(b) the testimony is based on sufficient facts or data [that reliably support the expert's opinion;]

(c) the testimony is the product of reliable [and objectively reasonable] principles and methods; and

(d) the expert has reliably applied the principles and methods to the facts of the case [and reached a conclusion without resort to unsupported speculation.]

***Reporter's Comments***

1. The brackets were suggested by the authors of the article that pointed out that courts are treating some 702 factors as questions of weight and not admissibility. I am not sure that any other necessary. For example, if the testimony is the product of reliable principles and methods, surely that would mean that the principles and methods are objectively reasonable. And if the facts or data are sufficient, that should mean that they reliably support the expert's opinion.



2. Another way to structure this is to put the preponderance language at the end as a hanging paragraph. Like this:

**Rule 702. Testimony by Expert Witnesses.**

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

- (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data [that reliably support the expert’s opinion;]
- (c) the testimony is the product of reliable [and objectively reasonable] principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case [and reached a conclusion without resort to unsupported speculation.]

The proponent must establish that all of the admissibility requirements of this rule are met by a preponderance of the evidence.

Restylists don’t like hanging paragraphs, but the virtue of this fix is that it does no damage to the rule that exists, and more importantly that it covers the question of qualifications as well as reliability. Under the previous alternative, the argument could be made that qualifications does not have to be established by preponderance, because the preponderance language refers to the “following” factors, and the qualification requirement precedes this language.

3. Another way to fix it so that all admissibility requirements would be covered is to restructure the rule to add qualifications to the lettered list:

**Rule 702. Testimony by Expert Witnesses.**

The following requirements must be established by a preponderance of the evidence for a witness to testify as an expert in the form of an opinion or otherwise: A witness who

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

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

(~~b~~) (c) the testimony is based on sufficient facts or data [that reliably support the witness's opinion;]

(~~e~~) (d) the testimony is the product of reliable [and objectively reasonable] principles and methods; and

(~~d~~) (e) the ~~expert~~ witness has reliably applied the principles and methods to the facts of the case [and reached a conclusion without resort to unsupported speculation.]

If this alternative is adopted, it might be better to put qualifications at the *end*, as a new subsection (e). This will be less disruptive to electronic searches.

Finally, if any of these drafting changes regarding admissibility and weight are adopted, the Advisory Committee Note will have to emphasize that no inference is being created as to other rules, i.e., by saying that the preponderance standard applies here, the Committee is not, by negative inference, saying that it doesn't apply in any other rule. The case will have to be made that the amendment was required because courts have ignored what is evident in the existing rule, and in the FRE as a whole --- that admissibility requirements other than conditional relevance are determined by a preponderance of the evidence.

## Conference materials for October 27

**From:** Daniel Capra <(b) (6)>  
**To:** Eric Lander <(b) (6)>, "Dror, Itiel" <(b) (6)>, Karen Kafadar <(b) (6)>, Tom Albright <(b) (6)>, "Isenberg, Alice R. (LD) (FBI)" <(b) (6), (b)(7)(C), (b)(7)(E) per FBI>, "Ballou, Susan M (Fed)" <(b) (6)>, Jeff Salyard <(b) (6)>, (b)(6) Alex Kozinski <(b)(6) Jed Rakoff <(b) (6)>, (b) (6)>, (b)(6) K Michael Moore <(b) (6)>, Ronald J Allen <(b) (6)>, "Kaye, David" <(b) (6)>, Jonathan J Koehler <(b) (6)>, (b)(6) Jane Moriarty, "Murphy, Erin" <(b) (6)>, Stephen Saltzburg <(b) (6)>, "Hunt, Ted (ODAG)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>, Chris Fabricant <(b) (6)>, Anne Goldbach <(b) (6)>, (b)(6) Patti Saris <(b)(6) Paul Grimm <(b)(6) Hafer, Zachary (USAMA)" <(b) (6)>, "Lightfoot, Lori" <(b) (6)>, (b)(6) Laura Shamp, Thomas Sobol <(b) (6)>  
**Cc:** (b)(6) Debra Livingston <(b)(6) David Campbell <(b)(6) Daniel Coquillette, (b)(6) Shelly Cox <(b)(6) Shapiro, Elizabeth (CIV)" <(b) (6)>  
**Date:** Wed, 11 Oct 2017 13:00:51 -0400  
**Attachments:** Rule 702 conference BC Law School conference materials.docx (42.14 kB)

Attached is a file representing the conference materials for October 27 --- Agenda, Speaker bios, and relevant rules for discussion

I would like to cover a couple of process issues

- 1 We will be operating under significant time constraints given the number of participants. And while the individual presentations are obviously critical, I believe that it is equally important to save time for discussion among the participants, and for questions from and discussion with members of the Advisory Committee. Therefore, I implore each of the speakers to adhere to the time limits that were originally stated --- no more than 10 minutes for the initial presentation. You can save what you don't get to for the general discussion. I will be giving two minute warnings. Thanks to everyone in advance for bringing their talks within the time limits.
2. While the PCAST report was the reason that the forensics panel was conceived, the Conference is not about the merits of the PCAST report. It is about what the current problems are in forensic expert testimony and whether the problems that do exist can be usefully regulated by the courts and, specifically, by rulemaking. I am hoping that the conversation will be about those matters, rather than a line by line attack or support on the PCAST report.
3. If you wish to do a powerpoint presentation, please bring it on a USB stick. BC prefers to have it provided to them in that way. It would probably be a good idea to get everything loaded before we start at 8:30.
4. The Conference is scheduled to conclude at 1 and lunch will follow. We start at 8:30 sharp.

If you have any questions or anything you need help with in terms of logistics or otherwise, please let me know.

Thank you all so much. We are really excited about your participation in this important event. See you all soon.

Daniel J. Capra  
Reed Professor of Law  
Fordham Law School  
New York, New York

(b) (6)

**Judicial Conference Advisory Committee on Evidence Rules  
Symposium on Forensic Expert Testimony, *Daubert*, and Rule 702**

**Boston College Law School**

**October 27, 2017**

**Conference Materials**

**I. Agenda**

***Introductory Remarks***

**Dean Vincent D. Rougeau**, Boston College Law School

**Hon. David Campbell**, Chair of the Committee on Rules of Practice and Procedure

**Hon. William K. Sessions, III**, former Chair of the Advisory Committee on Evidence Rules

***Panel One: Forensic Evidence***

**Scientists**

**Dr. Eric Lander**, President and founding director of the Broad Institute of MIT and Harvard; co-chair of the President's Council of Advisors on Science and Technology (PCAST).

***Topic: Rulemaking to Help Assure the Validity of Forensic Expert Testimony.***

**Dr. Itiel Dror**, University College London (UCL) and Cognitive Consultants International.

***Topic: "Reliability and Biasability of Expert Evidence"***

Expert evidence is often based on human perception, judgment, interpretation and decision making. These often include subjective elements. Subjectivity is not necessarily a bad thing, but it can introduce two major concerns. First, reliability (in the scientific sense



of consistency and reproducibility), that is, will different experts reach the same conclusions (the inter- between-expert reliability); and more basic, will the same expert, examining the same data, reach the same conclusions (the intra- within-expert reliability). The second concern is biasability, the biasing influence of irrelevant contextual information, as well as target driven bias (whereby the experts work 'backward' from the 'target' suspect to the evidence, rather than the evidence itself driving the forensic work). The Hierarchy of Expert Performance (HEP) demonstrates that expert evidence suffers from both issues of reliability and biasability, even in forensic fingerprint and mixture DNA evidence.

The problem is that forensic evidence is often misrepresented in court and is incorrectly regarded by most jurors (as well as judges, and the forensic experts themselves) as objective and impartial evidence. It is therefore important to make sure that there are minimal misconceptions about the true nature and weaknesses of forensic evidence. Furthermore, that the courts make sure that steps are taken by experts to deal with those weaknesses, such as LSU - Linear Sequential Unmasking (which stipulates that experts should only be exposed to relevant information and methods for ensuring experts work from the evidence to the suspect, not backwards). When expert evidence fails to meet these standards, it is biased and unreliable, and then it should be excluded. The fear of evidence being excluded will make a much needed positive impact on the way forensic work is carried out, resulting in evidence that is more impartial and reliable.

**Dr. Karen Kafadar**, Commonwealth Professor & Chair of Statistics at University of Virginia.

***Topic: Distinguishing Opinion and Relevance from Demonstrably Sufficient Science***

Rule 702 allows a witness to testify "in the form of an opinion or otherwise" if "the testimony is based on sufficient facts or data" and "is the product of reliable principles and methods" that have been "reliably applied". The determination of "sufficient" (facts or data), and whether the "reliable principles and methods" relate to the scientific question at hand, involve more discrimination than the current Rule 702 may suggest. Using examples from latent fingerprint matching and trace evidence (bullet lead and glass), Dr. Kafadar will offer some criteria that scientists often consider in assessing the "trustworthiness" of evidence, to enable courts to better distinguish between "trustworthy" and "questionable" evidence. The codification of such criteria may ultimately strengthen the current Rule 702 so courts can better distinguish between demonstrably scientific sufficiency and "opinion" based on inadequate (or inappurtenant) methods.

**Dr. Jeff Salyards**, Director of the Defense Forensic Science Center, Department of Defense .

***Topic: “Uncertainty, Error, and Mistake,” and the Difference Between “validation and Validation.”***

**Dr. Thomas Albright**, Professor and Conrad T. Prebys Chair, Salk Institute for Biological Studies.

***Topic: Why Eyewitnesses Fail***

Eyewitness identifications play an important role in the investigation and prosecution of crimes, but it is well known that eyewitnesses make mistakes, often with serious consequences. In light of these concerns, the National Academy of Sciences recently convened a panel of experts to undertake a comprehensive study of current practice and use of eyewitness testimony, with an eye towards understanding why identification errors occur and what can be done to prevent them. The work of this committee led to key findings and recommendations for reform, detailed in a consensus report entitled *Identifying the Culprit: Assessing Eyewitness Identification*. In this presentation, Dr. Albright will focus on the scientific issues that emerged from this study, along with brief discussions of how these issues led to specific recommendations for additional research, best practices for law enforcement, and use of eyewitness evidence by the courts.

**Dr. Alice R. Isenberg**, Deputy Assistant Director of the FBI Laboratory

***Topic: The Modern Practice of Forensic Science***

As a forensic practitioner, Dr. Isenberg will speak about forensic methods and actions to strengthen use of forensic science in the laboratory and in the courtroom. She will also discuss validation research in a federal laboratory and address the role and importance of quality assurance systems including, accreditation, testimonial training, and competency and proficiency testing.

**Susan Ballou**, Program Manager for the Forensic Sciences Research Program, National Institute of Standards and Technology (NIST).

***Topic: Getting The Science Right – Not The Focus of Rule of Evidence 702***

- Measurement science provides basis for testimony – data driven results required to justify position.

- Science is presented with increased specificity and certainty – supporting the selected principles and methods

## **Judiciary**

**Hon. Alex Kozinski**, Circuit Judge, Ninth Circuit Court of Appeals

*Topic: A Comment on the Science Presentations and the Role of Rule 702.*

**Hon. Jed S. Rakoff**, District Judge, Southern District of New York

*Topic: The Problem of Experts Overstating a “Match”*

**Hon. K. Michael Moore**, Chief Judge, Southern District of Florida

*Topic: The Need for a Flexible Rule*

Chief Judge Moore will be discussing the need for a flexible rule to enable trial court judges to assess the admissibility of expert opinions, especially as the legal landscape evolves. Specifically, Chief Judge Moore will address recent developments in drug prosecutions pertaining to synthetic drugs and assessing the reliability of experts in this area.

## Academics

**Professor Ronald J. Allen**, John Henry Wigmore Professor of Law, Northwestern Pritzker School of Law

***Topic: Fiddling While Rome Burns: the Story of the Federal Rules and Experts.***

Worrying about the “reliability” of some discipline with little assurance that it is has been applied correctly, and less assurance that the fact finder understands it, is to fiddle while Rome burns. This point derives from Professor Allen’s papers that explored the distinction between educational and deferential models of decision making.

**Professor David H. Kaye**, Distinguished Professor and Weiss Family Scholar, Penn State Law School

***Topic: Why Has Rule 702 Failed Forensic Science?***

Eight years ago, a committee of the National Academy of Sciences concluded that “[i]n a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem.” The committee also observed that “[f]ederal appellate courts have not with any consistency or clarity imposed standards ensuring the application of scientifically valid reasoning and reliable methodology in criminal cases involving *Daubert* questions.” This situation, it added, was “not surprising” given that *Daubert* is so “flexible.”

This presentation will elaborate on these conclusory remarks in four ways (time permitting). First, it will describe how ambiguities and flaws in the terminology adopted in *Daubert* combined with the opaqueness of forensic-science publications and standards have been exploited to shield some test methods from critical judicial analysis. Second, to promote an improved understanding of the necessary foundations for scientific and other expert testimony, it will sketch various meanings of the terms “validity” and “reliability” in science and statistics on the one hand, and in the rules and opinions on the admissibility of expert evidence, on the other. In this regard, it will skeptically consider the two-part definition of “validity” in a 2016 report of the President’s Council of Advisors on Science and Technology and will question the report’s effort to draw a bright line for the “validity” of pattern-matching testimony. Third, it will ask if the Federal Rules of Evidence should be revised to conform more closely to the usual scientific terminology. Finally, it will identify four ways to indicate uncertainty in forensic findings and will propose requiring statements about uncertainty when reporting outcomes of scientific tests.



**Professor Jonathan J. Koehler**, Beatrice Kuhn Professor of Law at Northwestern Pritzker School of Law

***Topic: Data and Forensic Science Opinions***

FRE 702 permits expert opinion testimony if the opinion is based on sufficient facts or data, if those facts or data are derived from reliable principles and methods, were reliably applied in the instant case, and if the opinion is helpful to the trier of fact. I will suggest that, in many instances, opinion testimony that is routinely provided by forensic scientists fails this test. The failure arises largely because trial judges have not required testifying forensic scientists in any area (including DNA) to provide meaningful data to help judges and jurors assess the probative value of forensic opinion testimony. Empirical evidence from studies with mock jurors hints at substantial confusion around this issue. I will suggest amending FRE 702 to ensure that forensic science opinion testimony – and other expert testimony that relies heavily on subjective human judgment – receives a more rigorous vetting at trial.

**Professor Jane Campbell Moriarty**, Carol Los Mansmann Chair in Faculty Scholarship, Duquesne University School of Law

***Topic: Judicial Gatekeeping of Forensic Science Feature-comparison Evidence.***

Courts generally admit feature-comparison evidence, despite little proof of scientific reliability. Why are courts generally unreceptive to challenges about the reliability of such evidence? It may be that judges (like most people) perceive feature-comparison evidence as fairly straightforward and intuitively accurate. This perception may cause courts to employ heuristic approaches to the evidence—that is, cognitive shortcuts that manage complexity—which can be influenced by common cognitive biases, such as belief perseverance and confirmation bias. By understanding that feature-comparison “matching” is a complex, multifaceted process, courts might engage in a deeper, science-based review to better analyze the shortcomings and limitations of such evidence.

**Professor Erin Murphy**, N.Y.U. Law School

***Topic: Machine-Generated Forensic Evidence***

Technology has dramatically changed the shape of evidence in criminal courts. Forensic comparisons increasingly rely on machine-generated information, such as the DNA match statistics produced by a probabilistic genotyping software program or the location data reported by a cell phone tracker. This talk probes whether rules designed for viva voce confrontation of isolated pieces of evidence require tweaking when applied to machine-generated evidence.

**Professor Stephen A. Saltzburg**, Wallace and Beverley Woodbury University Professor,  
George Washington University Law School

***Title: Requiring Appointment of a Defense Expert to Challenge the Government's Forensic Expert***

Professor Saltzburg will explore the question whether a defense lawyer confronting expert testimony and/or scientific tests by the government can provide effective assistance of counsel without having access to a defense expert to examine the government's forensics. The solution to the problem may be an amendment to Rule 706, or an appointment provision added to a new rule on forensic evidence.

## **Practitioners**

**Ted R. Hunt**, Senior Advisor on Forensics, United States Department of Justice

***Topic: The PCAST Report***

Mr. Hunt will speak directly to the PCAST report and offer the Department's official position on the report.

**Andrew Goldsmith**, Associate Deputy Attorney General and National Criminal Discovery Coordinator, United States Department of Justice

***Topic: The Reliability of the Adversarial System to Inform Factfinders About Any Genuine Issues as to the Reliability or Accuracy of Forensic Testimony.***

**Chris Fabricant**, Joseph Flom Special Counsel and Director of Strategic Ligation, The Innocence Project

***Topic: The 702 Requirement of Reliable Application***

Mr. Fabricant will discuss 702/*Daubert* as it relates to forensic sciences, with a particular focus on FRE 702(c)'s requirement that the testimony at issue be the product of reliable principles and methods, and how this requirement has been interpreted by courts in criminal cases.

**Anne Goldbach**, Forensic Services Director, Committee for Public Counsel Services, Public Defender Agency of Massachusetts.

***Topic: Rule 702(d) and Forensic Experts***

Ms. Goldbach will discuss Rule 702(d)'s requirement that expert testimony must demonstrate that the expert has reliably applied the principles and methods to the facts of the case, and how this requirement has been interpreted in criminal cases involving forensic experts in the First Circuit and Massachusetts courts.

***Concluding Remarks: Special Commentary by Professor Charles Fried, Beneficial Professor of Law, Harvard Law School.***

***Panel Two: Rule 702 and Daubert***

**Judiciary**

**Hon. Patti B. Saris**, Chief Judge, District of Massachusetts

***Topic: Daubert Gatekeeping and Complex Scientific Concepts***

Chief Judge Saris will address the challenges to courts in addressing *Daubert* motions where the scientific concepts are complex, like patent litigation or product liability. Her perspective is that *Daubert* does not have the liberalizing effect the Supreme Court anticipated but actually makes it harder to have expert evidence introduced. She will outline different approaches courts use to understand the science (like tutors).

**Hon. Jed S. Rakoff**, District Judge, Southern District of New York

***Topic: How Daubert is Working in Non-Forensic Cases, and How Trial Judges Seek to Avoid Daubert Rulings.***

**Hon. Paul W. Grimm**, District Judge, District of Maryland

***Topic: Structural Impediments for Judges Applying Rule 702 in Criminal Cases***

Courts encounter special difficulties in making reasoned *Daubert* rulings in criminal cases. Structural impediments include: 1) the speed at which criminal cases proceed; 2) the significantly less helpful criminal expert disclosure rules as compared with the civil rules disclosures; 3) the overlay of the plea bargaining process and pressure on defendants not to file motions; and 4) resource limits on the ability of public defenders and CJA panel counsel on hiring forensic experts. These limitations make it very difficult for trial judges to get the information they need to perform a *Daubert*/Rule 702 analysis sufficiently far in advance of trial.

## **Practitioners**

**Zachary Hafer**, Assistant U.S. Attorney, District of Massachusetts

***Title: Daubert from the Perspective of a Prosecutor***

Mr. Hafer will address Judge Grimm's remarks and speak further about the challenges of applying *Daubert* from the prosecutor's perspective.

**Lori Lightfoot**, Mayer Brown, Chicago

***Title: Making the Gatekeeping Function Meaningful***

Experience shows *Daubert* motions have become perfunctory, i.e. it is assumed that such motions will be filed, and not attacking an expert through a *Daubert* motion is the exception, not the rule --- which obviously is not the intent. Experience also indicates judges are very reluctant to grant a *Daubert* motion if there is even a colorable argument in support of the expert's proffered testimony. So, the challenge is how to have the rule serve as an appropriate gatekeeper without barring legitimate testimony, given the significant role that experts can play in a trial. Another issue is whether, and to what extent, the rulings on the *Daubert* motions influence the settlement decision.

**Laura M. Shamp**, Shamp Jordan Woodward, Atlanta

***Topic: Daubert from the Plaintiffs' Perspective***



Ms. Shamp will speak about the challenges that are faced by plaintiffs under *Daubert* and on whether an amendment to Rule 702 would be helpful to address those challenges.

**Thomas M. Sobol**, Hagens Berman, Boston

***Title: Problems in the Use of Expert Screening Tools***

Mr. Sobol will address two opposing forces in the use of *Daubert* and related expert screening tools. On the one hand, the perceived or actual overuse of these tools occasionally leads to a lack of focus to cull out those portions of expert testimony that truly ARE contrary to law or the relevant professional standards. On the other hand, these tools too often provide a vehicle for judicial intervention into the jury's fact finding role. The solution is more selective attacks by counsel, as opposed to shotgun motions.

---

***Concluding Remarks on the Conference***

**Hon. Debra Livingston**, Chair of the Advisory Committee on Evidence Rules.

## II. Speaker Bios

### **Dr. Thomas D. Albright**

Dr. Thomas D. Albright is Professor and Conrad T. Prebys Chair at the Salk Institute for Biological Studies in La Jolla, California. His laboratory seeks to understand the brain bases of visual perception, memory and visually-guided behavior. Albright received a Ph.D. in psychology and neuroscience from Princeton University. He is a member of the US National Academy of Sciences, a fellow of the American Academy of Arts and Sciences, and a fellow of the American Association for the Advancement of Science.

Albright served as co-chair of the US National Academy of Sciences Committee on Scientific Approaches to Eyewitness Identification, which produced the 2014 report *Identifying the Culprit: Assessing Eyewitness Identification*. He is a member of the US National Academy of Sciences Committee on Science, Technology, and Law, and serves on the US National Commission on Forensic Science.

### **Professor Ronald J. Allen**

Professor Allen is the John Henry Wigmore Professor of Law at Northwestern University, in Chicago, IL. He did his undergraduate work in mathematics at Marshall University and studied law at the University of Michigan. He is an internationally recognized expert in the fields of evidence, criminal procedure, and constitutional law. He has published seven books and over 100 articles in major law reviews. He has been quoted in national news outlets hundreds of times, and appears regularly on national broadcast media on matters ranging from constitutional law to criminal justice. He has worked with various groups in China to help formulate proposals for legal reform, and he was recently retained by the Tanzanian Government to assist in the reform of their evidence law. He is a member of the American Law Institute, has chaired the Evidence Section of the Association of American Law Schools, and was Vice-chair of the Rules of Procedure and Evidence Committee of the American Bar Association's Criminal Justice Section.

### **Susan Ballou**

Susan Ballou has been involved in NIST research for the past 17 years. She is the Program Manager for the Forensic Sciences Research Program within the Special Programs Office at the National Institute of Standards and Technology (NIST), Gaithersburg, MD. She is also the Federal Officer for the NIST Forensic Science Center of Excellence based at Iowa State University and appropriately titled: the Center for Statistics and Applications in Forensic Evidence (CSAFE). Prior to NIST, she served as the lead serologist for the Montgomery County Police Department (MCPD) Crime Laboratory in Rockville, Maryland. Several of her cases have been on the highly

acclaimed TV series, *Forensic Files*. Before the MCPD she worked for the Commonwealth of Virginia Division of Consolidated Laboratory Services at their Merrifield location where she conducted analysis on evidence suspected of containing illicit drugs, body fluids and hairs and fibers. Her expertise with the Virginia system grew from her prior position as chemist in the Connecticut Office of the Chief Medical Examiner under the supervision of Chief Toxicologist, Dr. Randall Baselt. She holds a Master of Science degree in Biotechnology from The Johns Hopkins University and a Criminal Justice Undergraduate degree from the University of New Haven, West Haven, Connecticut. Qualified as an Expert in 180 court cases she has ventured beyond the crime laboratory to assist with crime scene investigations and has taught this information at The Judge Advocate General's Legal School and Center in Charlottesville, Virginia. She has served on the ASTM E30 Forensic Science committee and held the position of chair receiving the prestigious ASTM International Award of Merit with the honorary title of Fellow from Committee E30. She currently is the President-Elect of the American Academy of Forensic Sciences (AAFS) a 7000 member strong association. She holds fellow status in the AAFS and received the AAFS Criminalistics Section Mary E. Cowan Outstanding Service Award. She has authored book chapters, scientific papers and participated in documentary standards development during her membership in several forensic science related scientific working groups.

### **Dr. Itiel Dror**

After finishing his Ph.D. in psychology at Harvard University, Itiel Dror pursued his interest in expert performance. Along with his theoretical laboratory based research he has conducted fieldwork with a variety of experts (such as with U.S. Air Force pilots, frontline police officers, forensic examiners, and medical professionals). Dr. Dror's research has demonstrated that specific components in the cognitive underpinning of expertise entail vulnerabilities. Building on these insights he developed unique ways to combat these weaknesses and improve expert performance. Dr. Dror has published over 100 articles and is on the editorial board of a variety of scientific journals (such as *Science & Justice*, *Pragmatics & Cognition*, and the *Journal of Applied Research in Memory & Cognition*). He has trained judges in a variety of countries (e.g., the United States, United Kingdom, and Taiwan), as well as many forensic experts in law enforcement agencies (e.g., the FBI, NYPD, San Francisco PD, Boston PD, & LAPD in the United States, and in other countries, such as the Netherlands, Finland, Canada, Brazil, Singapore, Taiwan, and Australia). Dr. Dror now divides his time between academic work at University College London (UCL) and applied work at Cognitive Consultants International (CCI-HQ). More information is available at: [www.cci-hq.com](http://www.cci-hq.com)

### **M. Chris Fabricant, Esq.**

As the Joseph Flom Special Counsel and Director of Strategic Litigation, M. Chris Fabricant leads the Innocence Project's Strategic Litigation Department, whose attorneys develop and execute national litigation strategies to address the leading causes of wrongful conviction,

including eyewitness misidentification, the misapplication of forensic sciences and false confessions. Previously, he was a clinical law professor and the director of the Criminal Justice Clinic at the Pace Law School, where he was named a "Bellows Scholar" by the Association of American Law Schools, Clinical Legal Education Section. Mr. Fabricant has over a decade of criminal defense experience at the state and federal, trial and appellate levels with The Bronx Defenders and Appellate Advocates.

### **Anne Goldbach, Esq.**

Anne Goldbach is the Forensic Services Director for the Committee for Public Counsel Services. After graduating from Boston College Law School, Ms. Goldbach joined the Massachusetts Defenders Committee as a public defender in 1978. After the creation of CPCS, she joined the staff of Roxbury Defenders in January, 1985, where she became a supervising attorney; she was selected as Attorney in Charge of the Boston office in November, 1987. After running the Boston Trials Unit for 10 years, she became CPCS' Director of Forensic Service in November of 1997. In that capacity, she acts as a resource on forensics issues and experts for public defenders and bar advocates across the state.

Throughout her career, Ms. Goldbach has been actively involved in continuing legal education and criminal defense training programs, and has lectured on numerous forensics topics. She has been a frequent lecturer, writer and moderator for Mass. Continuing Legal Education, CPCS conferences and training programs, as well as other CLE training programs. She has served on the Board of Directors of the Mass. Council for Public Justice. She serves on the board of the Thomas J. Drinan Memorial Fellowship Fund at Suffolk University Law School. She is a non-voting member of the state's Forensic Sciences Advisory Board. She is a past president and current board member of MACDL, Massachusetts Association of Criminal Defense Lawyers.

In May 2000, Ms. Goldbach received the Hon. David S. Nelson Public Interest Law Award from the Boston College Law School Alumni Association. In May 2013, Ms. Goldbach received the Edward J. Duggan Public Defender Award from CPCS for zealous advocacy and outstanding legal services. In April 2014, Boston College Law School's Women's Law Center gave her the annual "Woman of the Year" award and in June, 2016 she received the Clarence Gideon Award from the Massachusetts Association of Criminal Defense Lawyers.

### **Andrew D. Goldsmith, Esq.**

Mr. Goldsmith was appointed in January 2010 as the Justice Department's first National Criminal Discovery Coordinator. In this role, he oversees a wide range of national initiatives designed to provide federal prosecutors and other law enforcement officials with training and resources relating to criminal discovery, including electronic discovery. As Associate Deputy



Attorney General, he is also responsible for topics concerning professional responsibility, recording of custodial statements, legal education, and environmental matters. Mr. Goldsmith previously served as the First Assistant Chief of DOJ's Environmental Crimes Section, and successfully prosecuted the *Atlantic States* case in New Jersey during 2005-06, an eight-month trial that is the longest environmental crimes-related prosecution in U.S. history. His articles on criminal e-discovery have appeared in the *United States Attorneys' Bulletin*. In 2016, Mr. Goldsmith earned his fourth Attorney General's Award when he received the Claudia J. Flynn Award for Professional Responsibility in recognition of his efforts to ensure that department attorneys carry out their duties in accordance with the rules of professional conduct.

He previously served as an Assistant U.S. Attorney for the District of New Jersey. Mr. Goldsmith started out his legal career as an Assistant District Attorney in the Manhattan D.A.'s Office during the high crime era of the 1980's. Mr. Goldsmith graduated *cum laude* in 1983 from Albany Law School, which presented to him in 2008 its Distinguished Alumni in Government Award. He received his B.S. degree in biology in 1979 from Cornell University, which selected him in 2014 for inclusion on its list of Distinguished Classmates.

### **Hon. Paul W. Grimm**

Paul W. Grimm serves as a District Judge for the United States District Court for the District of Maryland. He sits at the Greenbelt, Maryland courthouse located near Washington D.C. He was appointed to the Court on December 10, 2012. Previously, he was appointed to the Court as a Magistrate Judge in February 1997 and served as Chief Magistrate Judge from 2006 through 2012. In September, 2009 the Chief Justice of the United States appointed Judge Grimm to serve as a member of the Advisory Committee for the Federal Rules of Civil Procedure where he served until September, 2015 as the chair of the Discovery Subcommittee. Judge Grimm is a member of the American Law Institute, and has been an adjunct professor of law at the University of Baltimore School of Law and the University of Maryland School of Law, where he taught courses on evidence and discovery, and he has written extensively on both topics. Judge Grimm received his BA from the University of California, Davis, his JD from the University of New Mexico, and his LLM from Duke University.

### **Zachary R. Hafer, Esq.**

Zachary R. Hafer has extensive experience leading the investigation and prosecution of high-profile federal criminal cases, including capital murder, public corruption, RICO, mail and wire fraud, money laundering, and drug trafficking. Most recently, he was the lead prosecutor in the four-month capital retrial *United States v. Gary Lee Sampson*. During the five-week defense case in *Sampson*, the prosecution cross-examined nearly 50 witnesses, including 12 experts in the fields of neuroimaging, neuropsychology, neuropsychiatry, forensic pathology, and statistical analysis of life expectancy. Mr. Hafer has briefed and argued several appeals in the First Circuit and has twice received the Attorney General's Award: (1) in 2010 for leading a years-long

international drug trafficking and money laundering investigation in which U.S. and Colombian law enforcement arrested 78 drug traffickers and seized approximately \$10 million in cash and thousands of kilograms of cocaine; and (2) in 2014 for his work as a trial AUSA in *United States v. James "Whitey" Bulger*. Mr. Hafer began his career as a law clerk for U.S. District Judge Shirley W. Kram in the Southern District of New York and was also in private practice at Debevoise & Plimpton in the firm's New York office prior to joining the Department of Justice in 2007. Mr. Hafer received a full-tuition, merit scholarship to the University of Virginia School of Law, from which he graduated in 2003. He graduated *cum laude* from Dartmouth College in 1999, with High Honors in English.

### **Ted R. Hunt, Esq.**

Ted R. Hunt is Senior Advisor to the Department of Justice on Forensic Science. Prior to his appointment by the Attorney General, he was Chief Trial Attorney at the Jackson County Prosecutor's Office in Kansas City, Missouri, where he served for 25 years as a state level prosecutor and managed a large staff of trial attorneys. During that time, Mr. Hunt prosecuted more than 100 felony jury trials, the vast majority of which involved the presentation of forensic evidence.

Mr. Hunt is a former member of the National Commission on Forensic Science, the ASCLD/LAB Board of Directors, the Missouri Crime Lab Review Commission, the OSAC Legal Resource Committee, and the NDAA DNA Advisory Group. He also served as a member of the International Association of Chiefs of Police (IACP) Forensic Science Committee, and was an Invited Guest on the Scientific Working Group on DNA Analysis Methods (SWGDM) Next Generation Sequencing Working Group.

### **Dr. Alice R. Isenberg**

Dr. Isenberg is the Deputy Assistant Director of the FBI Laboratory. She entered duty at FBI in 1998 and has previously served as the Section Chief of the Biometrics Analysis Section, Unit Chief of the Mitochondrial DNA Unit, and as a forensic examiner in the DNA unit. While Chief of the Biometrics section, she managed the elimination of an offender DNA backlog of over 300,000 samples and the casework DNA backlog involving over 2700 criminal cases. She also facilitated the deployment of a new version of Combined DNA Index System (CODIS) software and implementation of numerous automated DNA techniques as well as the development of Rapid DNA capability. Dr. Isenberg earned her Master of Science and Ph.D. in Analytical Chemistry from the University of Virginia.

## **Dr. Karen Kafadar**

Karen Kafadar is the Commonwealth Professor & Chair of Statistics at University of Virginia. She received her Ph.D. in Statistics from Princeton University, and previously held positions at NBS (now NIST), Hewlett Packard's RF/Microwave R&D Division, National Cancer Institute, University of Colorado-Denver, and Indiana University. Her research focuses on robust methods, exploratory data analysis, characterization of uncertainty in the physical, chemical, biological, and engineering sciences, and methodology for the analysis of screening trials. She served on the National Academy of Sciences' Committees that led to "Weighing Bullet Lead Evidence" (2004), "Strengthening the Forensic Science System in the United States: A Path Forward" (2009), "Review of the Scientific Approaches Used During the FBI's Investigation of the Anthrax Letters" (2011), "Evaluating Testing, Costs, and Benefits of Advanced Spectroscopic Portals" (2011), and "Identifying the Culprit: Assessing Eyewitness Reliability" (2014). She also served on the governing boards for ASA, IMS, ISI, and NISS, is a member of OSAC's FSSB, and chairs OSAC's Statistical Task Group and ASA's Advisory Committee on Statistics in Forensic Science. She is past Editor of *JASA Reviews* (1996-98) and *Technometrics* (1999-2001), is currently Health & Life Sciences Editor for *The Annals of Applied Statistics*, and is an Elected Fellow of the ASA, AAAS, and ISI.

## **Professor David H. Kaye**

David H. Kaye is Distinguished Professor and Weiss Family Scholar at Penn State Law, a member of the graduate faculty of Penn State University's Program in Forensic Science, and Regents' Professor Emeritus of Law and of Life Sciences at Arizona State University. He has held research or teaching positions at Cornell University, Duke University, the University of Chicago, the University of Virginia, and universities in England and China.

Professor Kaye was an Assistant Watergate Special Prosecutor, an associate in a private law firm in Portland, Oregon, and a law clerk to Judge Alfred T. Goodwin, U.S. Court of Appeals for the Ninth Circuit. He holds degrees in law (Yale University), astronomy (Harvard University), and physics (MIT).

Professor Kaye's research and teaching focuses on the law of evidence, statistics, criminal procedure, forensic science, and forensic genetics. His publications include textbooks on statistics and on scientific evidence; treatises on evidence and scientific evidence; and over 170 articles and letters in journals of law, philosophy, psychology, medicine, genetics, forensic science, and statistics. He is the author or a coauthor of *The Double Helix and the Law of Evidence* (Harvard University Press), the *Handbook of Forensic Statistics* (forthcoming), *McCormick on Evidence*, *The New Wigmore, Modern Scientific Evidence* (first four editions), and the Federal Judicial Center's Reference Manual on Scientific Evidence.

Professor Kaye has served on committees of the American Statistical Association, the National Academy of Sciences, the National Commission on Forensic Science, the National Commission on the Future of DNA Evidence, the National Institutes of Health, the National

Institute of Standards and Technology, the Organization of Scientific Area Committees for Forensic Science (OSAC), and the International Conferences on Forensic Inference and Statistics. He is a recipient of the OSAC Distinguished Service Award.

### **Professor Jonathan J. Koehler**

Jonathan “Jay” Koehler is the Beatrice Kuhn Professor of Law at Northwestern Pritzker School of Law. He has a B.A. from Pomona College (Philosophy), and an M.A. and PhD in Behavioral Sciences from the University of Chicago. His research focuses on issues in forensic science, decision theory, and juror decision making. He is an editor of *Law, Probability & Risk*, and a consulting editor of *Judgment and Decision Making*. Prior to joining Northwestern in 2010, he was a University Distinguished Teaching Professor at The University of Texas at Austin (business school), and a professor at Arizona State University (business and law schools).

### **Hon. Alex Kozinski**

Judge Kozinski was appointed United States Circuit Judge for the Ninth Circuit on November 7, 1985, and served as Chief Judge from 2007 to 2014. He graduated from UCLA, receiving an A.B. degree in 1972, and from UCLA Law School, receiving a J.D. degree in 1975.

Prior to his appointment to the appellate bench, Judge Kozinski served as Chief Judge of the United States Claims Court, 1982-85; Special Counsel, Merit Systems Protection Board, 1981-82; Assistant Counsel, Office of Counsel to the President, 1981; Deputy Legal Counsel, Officer of President-Elect Reagan, 1980-81; Attorney, Covington & Burling, 1979-81; Attorney, Forry Golbert Singer & Gelles, 1977-79; Law Clerk to Chief Justice Warren E. Burger, 1976-77; and Law Clerk to Circuit Judge Anthony M. Kennedy, 1975-76.

### **Dr. Eric Lander**

Eric Lander is president and founding director of the Broad Institute of MIT and Harvard. A geneticist, molecular biologist, and mathematician, he has played a pioneering role in the reading, understanding, and biomedical application of the human genome. He was a principal leader of the Human Genome Project.

With his colleagues, Lander has developed and applied powerful methods for discovering the molecular basis of rare genetic diseases, common diseases, and cancer. He has done pioneering work on human genetic variation; human population history; genome evolution; regulatory elements; long non-coding RNAs; three-dimensional folding of the human genome; and genome-wide screens to discover the genes essential for biological processes using CRISPR-based genome editing.



Lander is professor of biology at MIT and professor of systems biology at Harvard Medical School. From 2009 to 2017, he served as co-chair of the President's Council of Advisors on Science and Technology for President Barack Obama.

Lander's honors and awards include the MacArthur Fellowship, the Breakthrough Prize in Life Sciences, the Albany Prize in Medicine and Biological Research, the Gairdner Foundation International Award of Canada, the Dan David Prize of Israel, the Mendel Medal of the Genetics Society in the UK, the City of Medicine Award, the Abelson Prize from the AAAS, the Award for Public Understanding of Science and Technology from the AAAS, the Woodrow Wilson Prize for Public Service from Princeton University, and the James R. Killian Jr. Faculty Achievement Award from MIT.

### **Lori Lightfoot, Esq.**

Lori Lightfoot is a partner at Mayer Brown in Chicago. She is a trial attorney, investigator and risk manager. Both as a civil litigator and as Assistant US Attorney in the Criminal Division of the US Attorney's Office, Northern District of Illinois (1996–2002), Lori has tried over 20 federal and state jury and bench trials. She has also argued cases in state and federal appellate courts, and she has successfully conducted numerous internal investigations. From 2002 to 2005, Lori worked with the City of Chicago as Interim First Deputy Procurement Officer, Department of Procurement Services (DPS); General Counsel and Chief of Staff, Office of Emergency Management and Communications (OEMC); and Chief Administrator, Office of Professional Standards (OPS) of the Chicago Police Department. At OPS, Lori managed a 100-person office of civilian investigators charged with investigating police-involved shootings, allegations of excessive force and other misconduct alleged against Chicago police officers. She also coordinated joint investigations with state and federal criminal authorities and facilitated the implementation of new compliance and risk-management systems that included redesign of the disciplinary processes for sworn and civilian members, creation of a management intervention program for problem employees, and targeted tracking of litigation costs associated with complaints against department members. Lori has been associated with Mayer Brown since 2005 and, previously, between 1990 and 1996. Earlier, she served as Law Clerk to The Honorable Charles Levin, Michigan Supreme Court (1989–1990). She is a graduate of the University of Michigan and the University of Chicago Law School.

### **Hon. K. Michael Moore**

Chief Judge K. Michael Moore received his B.A. in Economics from Florida State University in 1972 and his J.D. from Fordham Law School in 1976. Judge Moore served as an Assistant United States Attorney for the Southern District of Florida from 1976 to 1981. From 1982 to 1986 he served as Assistant United States Attorney for the Northern District of Florida and held supervisory, Chief Assistant and Court-appointed United States Attorney positions.

In 1987, he received the first of three presidential appointments requiring United States Senate confirmation when President Ronald Reagan appointed Judge Moore to be United States Attorney for the Northern District of Florida. While United States Attorney, Judge Moore was also selected to serve on the Attorney General's Advisory Committee. As United States Attorney, Judge Moore was responsible for overseeing civil and criminal litigation on behalf of the United States for the northern third of the State of Florida.

In 1989, President George Bush appointed Judge Moore to be Director of the United States Marshals Service. In receiving this appointment, Judge Moore became the first presidentially appointed Director of our nation's oldest law enforcement agency. As Director, Judge Moore oversaw the Marshals Service's judicial security, witness security, fugitive apprehension, asset forfeiture, and prisoner transportation programs.

In 1992, President Bush appointed Judge Moore to the United States District Court for the Southern District of Florida. In July 2014, Judge Moore became the Chief Judge of the Southern District of Florida.

### **Professor Jane Campbell Moriarty**

Jane Campbell Moriarty is the Carol Los Mansmann Chair in Faculty Scholarship and Professor at Duquesne University School of Law in Pittsburgh, PA. She teaches Evidence, Scientific and Expert Evidence, Neuroscience and Law, and Professional Responsibility—all areas of her scholarship. Among her publications are a treatise, Giannelli, Imwinkelried, Roth & Moriarty, *Scientific Evidence* (Fifth Edition 2013, supps. 2014-2017) and a casebook, *Scientific and Expert Evidence* (Aspen, 2nd ed. 2011)(with John M. Conley) and several articles in the areas of science and law, judicial decision making, and legal ethics. Relevant articles include *Seeing Voices: Potential Neuroscience Contributions to a Reconstruction of Legal Insanity*, 85 *Fordham L. Rev.* 101 (2016); *The Legal and Policy Implications of Using Brain Imaging for Lie Detection*, 19 *Psych., Pub. Pol'y & Law* 222 (2013)(co-authored); "Will History Be Servitude?" *The NAS Report on Forensic Science and the Role of the Judiciary*, 2010 *Utah L. Review* 299 (2010); "Misconvictions" *Science and The Ministers of Justice*, 86 *Nebraska L. Rev.* 1 (2007); and *Forensic Science: Grand Goals, Tragic Flaws & Judicial Gatekeeping*, 44 *ABA Judges' Journal* 16 (2005)(with Michael Saks). She is currently working on a book for NYU Press entitled, *Are you Lying Now? Neurotechnology and Law* (2018).

### **Professor Erin E. Murphy**

Erin Murphy's research focuses on technology and forensic evidence in the criminal justice system. She is a nationally recognized expert in forensic DNA typing, and her work has been cited by multiple times by the Supreme Court. Her book, *Inside the Cell: The Dark Side of Forensic DNA* (Nation Books 2015), addresses the scientific, legal, and ethical challenges of forensic DNA typing. Murphy is also co-editor of the *Modern Scientific Evidence* treatise, presently serves as

the Associate Reporter for the American Law Institute's project to revise Article 213 of the Model Penal Code, and was elected to the ALI in 2013. She has shared her scholarly work with popular audiences through publications in *Scientific American*, *The New York Times*, *USA Today*, *Slate*, the *San Francisco Chronicle*, and the *Huffington Post*, and has offered commentary for numerous media outlets, including NPR, CNN, MSNBC, and NBC Nightly News.

A proud recipient of the Podell Distinguished Teaching Award in 2012, Murphy's course offerings include criminal law and procedure, evidence, forensic evidence, and professional responsibility in the criminal context. She joined the NYU faculty after five years at UC Berkeley School of Law. Prior to that, Murphy spent five years as an attorney with the Public Defender Service for the District of Columbia. She received her B.A. in comparative literature from Dartmouth College in 1995 and her J.D. from Harvard Law School in 1999, both magna cum laude. She clerked for Judge Merrick B. Garland on the U.S. Court of Appeals for the D.C. Circuit.

### **Hon. Jed S. Rakoff**

Jed S. Rakoff has served since March 1996 as a U.S. District Judge for the Southern District of New York. He also frequently sits by designation on the 2<sup>nd</sup> and 9<sup>th</sup> Circuit Courts of Appeals. Judge Rakoff holds the position of Adjunct Professor at Columbia Law School -- where he teaches courses in white collar crime, science and the law, class actions, and the interplay of civil and criminal law --and Adjunct Lecturer at Berkeley Law School. He has written over 145 published articles, 635 speeches, 1500 judicial opinions, and co-authored 5 books. He is also a regular contributor to the *New York Review of Books*.

Judge Rakoff holds a B.A. degree from Swarthmore College (1964), an M.Phil. degree from Oxford University (Balliol, 1966), and a J.D. degree from Harvard Law School (1969). Following law school, he clerked for the late Hon. Abraham L. Freedman, US Court of Appeals, Third Circuit, and was then an associate at the Debevoise law firm. From 1973-80, he served as an Assistant United States Attorney Office in the Southern District of New York, the last two years as Chief of Business & Securities Fraud Prosecutions. Thereafter, before going on the bench, he was a partner at two large law firms in New York, specializing in white collar criminal defense and civil RICO.

Judge Rakoff served on the National Commission on Forensic Science and as co-chair of the National Academy of Science's Committee on Eyewitness Identification. He served on the New York City Bar Association's Executive Committee and was chair of the Association's Honors and Criminal Law Committees. He was Chair of the Second Circuit's Bankruptcy Committee, and Chair of the Southern District of New York's Grievance Committee and Criminal Justice Advisory Board. He served on Swarthmore College's Board of Managers, on the Governance Board of the MacArthur Foundation's Project on Law and Neuroscience, and on the Committee on the Development of the Third Edition of the Manual on Scientific Evidence. He has assisted the U.S. Government in the training of foreign judges in Azerbaijan, Bahrain, Bosnia, Dubai, Iraq, Kuwait, Morocco, Saudi Arabia, and Turkey. He is a Member of the American Academy of Arts and Sciences and of the American Law Institute. He is a Judicial Fellow of the American College of

Trial Lawyers and the American Board of Criminal Lawyers. He was a Director of the New York Council of Defense Lawyers.

## **Dr. Jeff Salyards**

Dr. Jeff Salyards, PhD, MFS, is the Executive Director of the Defense Forensic Science Center, Forest Park, Georgia. Prior to his current position, he served as the Chief Scientist for the U.S. Army Criminal Investigation Laboratory. Before coming to USACIL, he was a Principal Analyst with Analytical Services and authored a study about the best methods to train military operators in material collection during the conduct of operations.

He holds a PhD in Chemistry from Montana State University, a Masters of Forensic Sciences from The George Washington University and has completed a Fellowship in Forensic Medicine from the Armed Forces Institute of Pathology.

A former Director of the Defense Computer Forensic Laboratory and AFOSI Special Agent, he has 26 years of combined experience in investigations, forensic consulting and teaching. He served as the Deputy for Operations and Assistant Professor at the Air Force Academy Chemistry Department and was honored with the Outstanding Academy Educator Award. Dr. Salyards has served on the Board of Directors for the American Society of Crime Laboratory Directors/Laboratory Accreditation Board, the Department of Justice National Steering Committee for Regional Computer Forensic Laboratories, the Council of Federal Forensic Laboratory Directors, the ASCLD Board of Directors, and as a Commissioner for the Forensic Education Programs Accreditation Commission; he is a current member of the National Commission on Forensic Science. Dr. Salyards is a Fellow of the American Academy of Forensic Sciences. He is also a retired commissioned officer in the United States Air Force.

## **Professor Stephen A. Saltzburg**

Stephen A. Saltzburg has taught at The George Washington University Law School since 1990. In January 2004, he was named the Wallace and Beverley Woodbury University Professor. From 1990-2004, he was the Howrey Professor of Trial Advocacy, Litigation and Professional Responsibility. Professor Saltzburg founded and became the Director of the Masters Program in Litigation and Dispute Resolution in 1996. Before moving to George Washington, Professor Saltzburg taught at the University of Virginia School of Law from 1972 to 1990. He was named the first Chairholder of the Class of 1962 Endowed Chair. He co-founded the University of Virginia Law School Trial Advocacy Institute in 1981, which is now the National Trial Advocacy College at the University of Virginia Law School. He continues to be the Director of the College.

Professor Saltzburg served as Reporter for and then as a member of the Advisory Committee on the Federal Rules of Criminal Procedure and as a member of the Advisory Committee on the Federal Rules of Evidence. He was the Reporter for the Civil Justice Reform Act Committee for the District of Columbia District Court and then became Chair of that



Committee. From 1987 to 1988, Professor Saltzburg served as Associate Independent Counsel in the Iran-Contra investigation. In 1988 and 1989, Professor Saltzburg served as Deputy Assistant Attorney General in the Criminal Division of the Department of Justice, and in 1989 and 1990 was the Attorney General's ex officio representative on the United States Sentencing Commission. In June, 1994, the Secretary of the Treasury appointed Professor Saltzburg as the Director of the Tax Refund Fraud Task Force, a position he held until January, 1995. Professor Saltzburg is the author of numerous books and articles on criminal law and procedure, evidence, litigation and trial advocacy. He is a member of the ABA House of Delegates from the Criminal Justice Section (which he served as Chair) and the ABA Task Force on Cyber Security.

### **Hon. Patti B. Saris**

United States District Judge Patti B. Saris became Chief Judge of the United States District Court for the District of Massachusetts on January 1, 2013. She was Chair of the United States Sentencing Commission in Washington, DC from January, 2011 to January, 2017. She is a graduate of Radcliffe College '73 (Magna Cum Laude, Phi Beta Kappa) and Harvard Law School '76 (Cum Laude). After graduating from law school, she clerked for the Supreme Judicial Court, and then went into private practice. When Senator Edward M. Kennedy became chairman of the Senate Judiciary Committee, she moved to Washington D.C. and worked as staff counsel. She later became an Assistant United States Attorney, and eventually chief of the Civil Division. In 1986, Judge Saris became a United States Magistrate Judge, and in 1989, she was appointed as an Associate Justice of the Massachusetts Superior Court. In 1994, she was appointed to the United States District Court.

### **Laura M. Shamp, Esq.**

Laura is a plaintiff's trial lawyer practicing principally in the areas of medical negligence, product liability and catastrophic injury in both state and federal court. Laura graduated with honors from the Harvard Law School and thereafter clerked for the Honorable Robert H. Hall at the United States District Court for the Northern District of Georgia. In 1996 Laura formed her own firm where she focused on complex commercial litigation, medical negligence and product liability, almost exclusively on behalf of plaintiffs. In 2004, she returned to Harvard to study patient safety and medical errors under Don Berwick and Lucian Leape, leaders in the field of patient safety, and in 2005 obtained a Masters Degree in Public Health from the Harvard School of Public Health. Laura is dedicated to working to try and improve the quality of medical care in Georgia and serves on committees that review legislation to try and address issues of medical error and patient safety.

Laura is also a frequent appellate advocate. Her reported cases have helped shaped the law in Georgia regarding medical malpractice and the use of expert witness testimony in professional negligence cases. In February of 2015 Laura, along with her partners formed, a plaintiff's boutique firm --- Shamp Jordan Woodward --- focusing on complex civil trial work. Laura has received

seven and eight figure verdicts on behalf of her clients in difficult venues and with difficult cases in the past several years. Most recently she led a trial team in an *Engle* Progeny tobacco case in Florida, securing an \$11 million verdict for her client.

### **Thomas Sobol, Esq.**

Thomas M. Sobol has been the Managing Partner of Hagens Berman Sobol Shapiro's Boston office for fifteen years. He has almost thirty-five years of experience in complex civil litigation. Mr. Sobol currently leads drug pricing litigation seeking to recover overcharges for individuals, health plans, state governments, and others that pay for brand name and generic drugs. Mr. Sobol has been a lead negotiator in court-approved settlements with pharmaceutical companies totaling well over one billion dollars. He currently is court-appointed lead or co-lead counsel in *In re Solodyn Antitrust Litigation*, *In re Celebrex Antitrust Litigation*, *In re Lipitor Antitrust Litigation*, *In re Effexor Antitrust Litigation*, and other matters. Mr. Sobol was appointed lead counsel in *In re New England Compounding Pharmacy Litigation Multidistrict Litigation* MDL, representing more than 700 victims who contracted fungal meningitis or suffered other serious health problems caused by contaminated products produced by NECC. To date, related settlements exceed \$200 million. Mr. Sobol was also co-lead trial counsel in the Neurontin MDL, where the jury returned a \$142 million racketeering (RICO) verdict against Pfizer.

In the 1990s, Mr. Sobol served as Special Assistant Attorney General for the Commonwealth of Massachusetts and the states of New Hampshire and Rhode Island, and served as one of the private counsel for Massachusetts and New Hampshire in ground-breaking litigation against the tobacco industry. These cases led to significant injunctive relief and to monetary recovery in excess of \$10 billion to those states. Mr. Sobol practiced at the Boston firm of Brown Rudnick for about seventeen years, where he was a litigation partner for a decade.

Mr. Sobol served as judicial clerk for then-Chief Justice Allan M. Hale of the Massachusetts Appeals Court from 1983 to 1984. Mr. Sobol is a member of the bar of Massachusetts and has been appointed pro hac vice in numerous federal courts across the country. He graduated *summa cum laude* from Clark University in Worcester, Massachusetts in 1980 and was elected to Phi Beta Kappa in 1979. Mr. Sobol graduated *cum laude* from Boston University School of Law in 1983.

### III. Evidence Rules for the Discussion:

#### A. Rule 702:

<p><b>Rule 702. Testimony by Expert Witnesses</b></p>
<p>A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:</p> <ul style="list-style-type: none"><li>(a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;</li><li>(b) the testimony is based on sufficient facts or data;</li><li>(c) the testimony is the product of reliable principles and methods; and</li><li>(d) the expert has reliably applied the principles and methods to the facts of the case.</li></ul>

#### B. A Rule 702 “tweak” that might allow for a Committee Note on forensic expert testimony.

##### Rule 702. Testimony by Expert Witnesses

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

- (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data;
- (c) the testimony is the product of reliable principles, and methods, and empirical data; and
- (d) the expert has reliably applied the principles and methods to the facts of the case.

### C. A freestanding rule on forensic expert testimony:

**Rule 707. Testimony by Forensic Expert Witnesses.** If a witness is testifying on the basis of a forensic examination [conducted to determine whether an evidentiary sample is similar or identical to a source sample], [or: “testifying to a forensic identification”] the proponent must prove the following in addition to satisfying the requirements of Rule 702:

(a) the witness’s method is repeatable, reproducible, and accurate for its intended use --- as shown by empirical studies conducted under conditions appropriate to that use;

(b) the witness is capable of applying the method reliably --- as shown by adequate empirical demonstration of proficiency --- and actually did so; and

(c) the witness accurately states, on the basis of adequate empirical evidence, the probative value of [the meaning of] any similarity or match between the evidentiary sample and the source sample.

[future subdivisions might be added to codify specific forms of comparison such as ballistics. Or they might be added in separately numbered rules.]