
VIRGINIA LAW REVIEW

VOLUME 95

MARCH 2009

NUMBER 1

ARTICLES

INVALID FORENSIC SCIENCE TESTIMONY AND WRONGFUL CONVICTIONS

Brandon L. Garrett and Peter J. Neufeld***

THIS is the first study to explore the forensic science testimony by prosecution experts in the trials of innocent persons, all convicted of serious crimes, who were later exonerated by post-conviction DNA testing. Trial transcripts were sought for all 156 exonerees identified as having trial testimony by forensic analysts, of which 137 were located and reviewed. These trials most commonly included testimony concerning serological analysis and microscopic hair comparison, but some in-

* Associate Professor, University of Virginia School of Law.

** Co-Founder and Co-Director, The Innocence Project.

For their invaluable comments, we thank Kerry Abrams, Edward Blake, John Butler, Paul Chevigny, Simon Cole, Madeline deLone, Jeff Fagan, Stephen Fienberg, Samuel Gross, Eric Lander, David Kaye, Richard Lewontin, Greg Mitchell, John Monahan, Erin Murphy, Sinead O'Doherty, George Rutherglen, Stephen Schulhofer, William Thompson, Larry Walker, and participants at the Harvard Criminal Justice Roundtable, a UVA School of Law summer workshop, and the Ninth Annual Buck Colbert Franklin Memorial Civil Rights Lecture at the University of Tulsa College of Law. The authors thank the participants at the Fourth Meeting of the National Academy of Science, Committee on Identifying the Needs of the Forensic Sciences Community for their useful comments, and we thank the Committee for inviting our participation at that meeting. Preliminary study data were presented as a report to the Committee on February 1, 2008. We give special thanks to all of the attorneys who graciously volunteered their time to help locate and copy exonerees' criminal trial transcripts, and particularly Winston & Strawn, LLP for its work in locating and scanning files from DNA exonerees' cases. The fine research assistants whose contributions made this study possible are: Josephine Liu, T.J. Parnham, Kerry Shapleigh, and Jason Shyung. For excellent reference library assistance, we thank Jon Ashley, Michelle Morris and Kent Olson. Finally, we thank the Olin Program at University of Virginia Law School for its research support.

cluded bite mark, shoe print, soil, fiber, and fingerprint comparisons, and several included DNA testing. This study found that in the bulk of these trials of innocent defendants—82 cases or 60%—forensic analysts called by the prosecution provided invalid testimony at trial—that is, testimony with conclusions misstating empirical data or wholly unsupported by empirical data. This was not the testimony of a mere handful of analysts: this set of trials included invalid testimony by 72 forensic analysts called by the prosecution and employed by 52 laboratories, practices, or hospitals from 25 states. Unfortunately, the adversarial process largely failed to police this invalid testimony. Defense counsel rarely cross-examined analysts concerning invalid testimony and rarely obtained experts of their own. In the few cases in which invalid forensic science was challenged, judges seldom provided relief. This evidence supports efforts to create scientific oversight mechanisms for reviewing forensic testimony and to develop clear scientific standards for written reports and testimony. The scientific community can through an official government entity promulgate standards to ensure the valid presentation of forensic science in criminal cases and thus the integrity and fairness of the criminal process.

INTRODUCTION.....	3
I. SUMMARY OF FINDINGS, METHODOLOGY, AND PRINCIPLES	12
A. <i>The Study Set and Summary of Findings</i>	12
B. <i>Study Protocol and Types of Invalid Testimony Identified</i>	16
1. <i>Non-Probative Evidence Presented as Probative</i>	16
2. <i>Exculpatory Evidence Discounted</i>	18
3. <i>Inaccurate Frequency or Statistic Presented</i>	18
4. <i>Statistic Provided Without Empirical Support</i>	18
5. <i>Non-numerical Statements Provided Without Empirical Support</i>	19
6. <i>Conclusion that Evidence Originated from Defendant</i>	20
C. <i>Questioning the Incidence of Invalid Forensic Testimony</i>	23
D. <i>Ethics and Forensic Science Testimony</i>	30
E. <i>Legal Regulation of Forensic Testimony</i>	32
II. RESULTS: INVALID FORENSIC SCIENCE TESTIMONY	34
A. <i>Invalid Forensic Serology Testimony</i>	35

2009]	<i>Invalid Forensic Science Testimony</i>	3
	1. <i>Ignoring the Problem of Masking and Quantification</i>	36
	2. <i>Invalid Analysis of Presumptively Exculpatory Serology</i>	42
	3. <i>Additional Invalid Use of Population Statistics</i>	46
	B. <i>Invalid Microscopic Hair Comparison Testimony</i>	47
	1. <i>Invalid Probability Testimony</i>	51
	2. <i>Exaggeration of Probative Value of Hair Association Without Using Numbers</i>	55
	C. <i>Invalid DNA Testimony</i>	63
	D. <i>Invalid Bite Mark Testimony</i>	67
	E. <i>Additional Forensic Techniques</i>	71
	1. <i>Shoe Print Analysis</i>	71
	2. <i>Fingerprint Analysis</i>	72
	3. <i>Voice Analysis</i>	74
	F. <i>Forensic Science Misconduct Beyond Trial Testimony</i>	76
	1. <i>Withholding Forensic Evidence</i>	76
	2. <i>Gross Error in Analysis</i>	79
	3. <i>Failures to Conduct Elimination Testing or Comparison</i>	81
III.	REFORMING FORENSIC SCIENCE	84
	A. <i>The Roles of Judges, Prosecutors, and Defense Attorneys</i>	84
	1. <i>Invalid Prosecution Use of Forensic Science</i>	84
	2. <i>Failures of Defense Counsel</i>	89
	3. <i>Judicial Rulings on Forensic Science</i>	90
	B. <i>A National Framework for Reform of Forensic Science</i>	93
	CONCLUSION	96

INTRODUCTION

Over the past two decades, DNA testing technology has both enhanced and eroded the status of forensic science in criminal cases. Traditional forensic disciplines were unable to identify a perpetrator with any great discrimination. For example, conventional serology analysis of blood group substances was widely used in sexual assault cases through the 1980s. The underlying method was sound and frequencies of the A, B, and O blood types were derived from well-established and scientifically valid databases. While serology could exclude or place an individual within a per-

centage of the population with a given blood type, it could not distinguish particular individuals with any greater specificity.

Forensic science had advanced dramatically by 1989, when Gary Dotson became the first innocent person in the United States exonerated by post-conviction DNA testing. A jury convicted Dotson in 1979 of rape, and he was sentenced to 25–50 years in prison.¹ In 1988, DNA testing was conducted after the Governor of Illinois had denied Dotson a pardon, despite the victim's recantation in which she stated that she had fabricated her accusation to conceal consensual intercourse with her boyfriend.² Edward Blake, who pioneered the forensic application of the polymerase chain reaction ("PCR") technology, conducted the testing. He found that the DNA results excluded Dotson as the source for the male genetic profile, but that the victim's boyfriend was included.³ Based on those findings, Dotson's conviction was vacated.⁴

Blake also found that the State's forensic analyst's testimony at Dotson's trial was misleading. The analyst had testified that both Dotson and the semen donor possessed the B blood type, a type shared by only eleven percent of Caucasians. The problem was not with the methods used in the laboratory but with the testimony in the courtroom. While on the witness stand, the analyst did not tell the jury that the victim was also Type B and that her fluids were mixed in the sample. The Type B substances observed in the sample could have come entirely from the victim. Her genetic markers could have overwhelmed, or "masked," those from the semen; as Blake put it, "no genetic information was obtained about the semen donor."⁵ Thus, based on the testing methods available at the time, any male could have been the donor. It was misleading to suggest to the jury that a subset (11%) of the population including

¹ Dotson's accuser later wrote a confessional book explaining her false testimony titled *Forgive Me*. Cathleen C. Webb & Marie Chapien, *Forgive Me* (1985).

² See Edward Connors, Thomas Lundregan, Neal Miller & Tom McEwan, Nat'l Inst. of Justice, *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial* 51–52 (1996), available at http://www.dna.gov/case_studies/convicted_exonerated/dotson [hereinafter *Convicted by Juries, Exonerated by Science*].

³ Affidavit of Edward T. Blake at 23, D. Crim., *People of the State of Illinois v. Gary E. Dotson*, No. P.C. 4333 (July 29, 1985).

⁴ *Convicted by Juries, Exonerated by Science*, supra note 2.

⁵ Affidavit of Edward T. Blake, *State v. Gary E. Dotson*, supra note 3, at 13.

Dotson could have been the donor and imply that 89% of the population was excluded.⁶

Thus, scientific advances led to Dotson's exoneration, but invalid forensic science testimony had also supported his conviction. Two hundred thirty-two innocent persons have now been exonerated by post-conviction DNA testing.⁷ Several of those exonerations, like Dotson's, have triggered scrutiny of the use of forensic science.⁸ Scandals involving faulty work at some of our nation's preeminent crime laboratories, including several arising from exoneration cases, have led to investigations, audits, and efforts to provide independent oversight.⁹ At the same time, scientists, legislators, and lawyers have raised questions concerning the validity and reliability of certain forensic science techniques. The American Bar Association issued a set of reform principles,¹⁰ and courts increasingly scrutinize forensic evidence in criminal cases.¹¹ Such efforts, unlike this study, chiefly focus on either the reliability of forensic science techniques or whether the underlying methodology is sound.

⁶ Id. Blake, who generously offered comments on this paper, conducted post-conviction DNA analysis in several other exonerees' cases. Blake published extensively on conventional serology prior to his groundbreaking DNA work. See Curriculum Vitae, Edward T. Blake, http://www.fsalab.com/etb_cv.htm#publications.

⁷ See The Innocence Project Home Page, <http://www.innocenceproject.org> (providing count of U.S. post-conviction DNA exonerations; the number as of January 29, 2009, is 232). The first author of this Article conducted a study of how the first 200 DNA exonerees fared during appeals and post-conviction proceedings. See Brandon L. Garrett, Judging Innocence, 108 Colum. L. Rev. 55, 58–59 (2008).

⁸ See Robert Bazell, DNA Acquittals Shaking Up Forensic Science, NBC News, Feb. 12, 2008, <http://www.msnbc.msn.com/id/23113417>; Jane Campbell Moriarty, "Misconvictions," Science, and the Ministers of Justice, 86 Neb. L. Rev. 1 (2007).

⁹ See Margaret A. Berger, The Impact of DNA Exonerations on the Criminal Justice System, 34 J.L. Med. & Ethics 320, 321–22 (2006) (noting investigations of "mistakes due to the incompetence or fraud of particular analysts," some of which "have gone on for years"); Paul C. Giannelli, Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs, 86 N.C. L. Rev. 163 (2007) (describing news reports of scandals and oversight efforts); Maurice Possley et al., Scandal Touches Even Elite Labs: Flawed Work, Resistance to Scrutiny Seen Across U.S., Chi. Trib., Oct. 21, 2004, § 1, at 1.

¹⁰ Andrew E. Taslitz, Convicting the Guilty, Acquitting the Innocent: The ABA Takes a Stand, 19 Crim. Just. 18–19 (2005).

¹¹ See, e.g., *House v. Bell*, 386 F.3d 668, 708 (6th Cir. 2004) (Merritt, J., dissenting) ("High on the list of the causes for mistakes are the kinds of errors we see in this case: the misinterpretation or abuse of scientific evidence . . ."); *United States v. Bentham*, 414 F. Supp. 2d 472, 473 (S.D.N.Y. 2006) ("False positives—that is, inaccurate incriminating test results—are endemic to much of what passes for 'forensic science.'").

Meanwhile, Congress tasked the National Academy of Sciences (“NAS”) with examining ways to improve the quality of forensic sciences.¹² The Committee’s landmark report emphasized that a wide range of forensic disciplines lack validity, where “[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.”¹³ The NAS report noted that “[n]ew doubts about the accuracy of some forensic science practices have intensified with the growing numbers of exonerations resulting from DNA analysis (and the concomitant realization that guilty parties sometimes walk free).”¹⁴ The report recommended wholesale reforms to improve not just the reliability and accuracy of forensic science, but also its presentation, including the creation of an independent federal agency—a “National Institute of Forensic Science”—to establish and enforce the use of “standard terminology” for report writing and testimony.¹⁵ Those latter recommendations are important—and the trials of the exonerated show why.

This study raises a neglected question: even assuming that a particular forensic technique has been validated and deemed reliable for casework, how do we ensure that the data will be interpreted, reported, and testified to within appropriate scientific parameters? Traditionally, there has been almost no oversight of what scientists *say* in the courtroom once the court deems the method used valid and reliable. To look at the problem of forensic science testimony in the courtroom, this Article will examine for the first time a set of criminal trial transcripts in the cases of DNA exonerees. The study asks whether forensic science testimony in exonerees’ trials com-

¹² See Comm. on Identifying the Needs of the Forensic Science Cmty., Nat’l Research Council of the Nat’l Acads., *Strengthening Forensic Science in the United States: A Path Forward* (2009), available at http://www.nap.edu/catalog.php?record_id=12589 [hereinafter *Strengthening Forensic Science*]; see also Comm. on Sci., Tech., and Law, *The Nat’l Acads., Identifying the Needs of the Forensic Sciences Community*, available at <http://www8.nationalacademies.org/cp/projectview.aspx?key=48741>. That Committee asked the authors to examine the incidence of invalid forensic science testimony in trials of persons exonerated by post-conviction DNA testing.

¹³ *Strengthening Forensic Science*, supra note 12, at S-5.

¹⁴ *Id.* at 1-2.

¹⁵ *Id.* at S-14–S-19, 6-3–6-5, 7-19.

ported with valid scientific principles.¹⁶ Throughout this Article, “invalid” testimony denotes a conclusion not supported by empirical data.¹⁷ This study does not examine reliability—that is, whether a forensic methodology produces consistent results.¹⁸ Nor does it examine whether in a particular case, an examiner made a mistake or engaged in misconduct in the laboratory. Instead, this study ex-

¹⁶ The use of forensic science in criminal trials more generally remains little studied. Simon Cole, *Where the Rubber Meets the Road: Thinking About Expert Evidence as Expert Testimony*, 52 *Vill. L. Rev.* 803, 819–24 (2007) (describing the many challenges of conducting such research, particularly where “American trial practice” is “incompletely and sporadically reported”). Several reports note particular examples of forensic science that supported wrongful convictions. The first such study was *Convicted by Juries, Exonerated by Science*, supra note 2, issued by the National Institute for Justice (“NIJ”) in 1996, which examined 28 DNA exonerations. *Id.* at xii. That study did not address whether in the 28 cases forensic science was presented in a valid manner. The second author co-authored a book that described several cases involving use of “junk science,” but not how many of the first 86 DNA exonerations involved invalid forensic science. See Barry Scheck, Peter Neufeld & Jim Dwyer, *Actual Innocence* 158–71 (2000). Others discuss examples of invalid use of forensic science. See Samuel R. Gross et al., *Exonerations in the United States: 1989 Through 2003*, 95 *J. Crim. L. & Criminology* 523, 543 (2005) (stating that in 24 exoneration cases, including non-DNA cases, a forensic expert committed perjury). The first author identified cases among the first 200 exonerated by post-conviction DNA testing in which forensic science supported the conviction and then studied appellate and post-conviction proceedings challenging that evidence. See Garrett, supra note 7, §§ II.A–II.B.

¹⁷ *Daubert v. Merrill Dow Pharmaceutical*, 509 U.S. 579, 590 n.9 (1993) (“[S]cientists typically distinguish between ‘validity’ (does the principle support what it purports to show?) and ‘reliability’ (does application of the principle produce consistent results?).”). This study does not examine the reliability or validity of methods used, but rather the validity of conclusions reached by forensic scientists. Similarly, the *Reference Manual on Scientific Evidence* defines a valid measurement as one that “measures what it is supposed to.” See David H. Kaye & David A. Freeman, *Reference Guide on Statistics*, in *Reference Manual on Scientific Evidence* 103 (2d ed. 2000). John Monahan and Laurens Walker define validity in the context of inferences drawn from research projects as “whether the methods and analyses employed were sound enough to justify the inferences drawn by the researcher.” John Monahan & Laurens Walker, *Social Science in Law* 60 (6th ed. 2006). This Article does not refer to any of the more specialized uses of validity in the sciences, such as internal or external validity of research. *Id.* at 60. The word “invalid” is used not only because of its accepted meaning in the sciences, but also because this Article examines only whether testimony was supported by data, and does not in any way characterize the ethics or the state of mind of the analysts who testified.

¹⁸ This Article does not examine the reliability of a particular discipline or field or the validity of forensic science methods. See Samuel R. Gross & Jennifer L. Mnookin, *Expert Information and Expert Evidence: A Preliminary Taxonomy*, 34 *Seton Hall L. Rev.* 141, 146–47 (2003) (describing distinction between field and method validity). For discussion of critics and defense of such disciplines, see *infra* Section II.D–E.

amines the validity of testimony—that is, whether what analysts said in court was supported by empirical data.

Examining forensic science testimony in the cases of DNA exonerees has several important limitations. Of the persons exonerated by post-conviction DNA testing, 156 had testimony concerning forensic evidence at their criminal trials. One advantage of looking at these cases is that relevant trial materials could be readily obtained. Most exonerees had forensic science testimony in their cases, because almost all were convicted at trial, and most of the cases involved rapes for which there was preserved crime scene evidence that could later be tested for DNA. Many also had post-conviction lawyers seek DNA testing who retained copies of the trial records. However, the same features that made this set an attractive subject for study also make this set unrepresentative of typical criminal cases. The data set consists entirely of erroneous outcomes, or innocent people convicted at trial. In addition, most exonerees were convicted of rape, since in such cases DNA evidence can often be highly probative to the issue of identity. Very few criminal defendants are convicted at a trial, where most plead guilty, and fewer are convicted of felony rape.¹⁹ Most exonerees were also convicted in the 1980s, before DNA testing was common.

As a result, one cannot determine from these data whether invalid forensic science testimony was common in the past two decades or is today. These data cannot provide information about forensic testimony in other types of far more common criminal cases. Invalid forensic science testimony in wrongful conviction cases might be the tip of a much larger iceberg, but it also might not. To answer that question, a broader inquiry into testimony in other types of cases and current cases is necessary. Such an inquiry, though desirable, faces practical difficulties, as no entity systematically collects or examines forensic science testimony in criminal cases. The purpose here, having obtained data from this group of innocent convicts, is simply to describe the testimony in these trials. That testimony provides examples suggesting a worrisome problem. Without reaching any conclusions about the size of the problem, these data point to the need to further investigate the content of forensic sci-

¹⁹ See Garrett, *supra* note 7, § I.A.

ence testimony, particularly where the conclusions expressed by forensic scientists on the stand are largely unregulated.

Trial transcripts were obtained for 137 of the 156 exonerees identified as having testimony by forensic analysts called by the prosecution at their trials.²⁰ This study observed invalid forensic science testimony in the bulk of these trials. In 82 cases, or 60%, forensic analysts called by the prosecution provided invalid testimony. This invalid testimony chiefly involved serological analysis and microscopic hair comparison, but also other forensic techniques, such as bite mark, shoe print, and fingerprint comparisons. Three additional cases involved withholding of exculpatory forensic evidence. Moreover, the invalid testimony was not the product of just a few analysts in a few states, but of 72 forensic analysts employed by 52 laboratories or medical practices in 25 states.

Two basic types of invalid science testimony occurred in these cases: (1) the misuse of empirical population data, and (2) conclusions regarding the probative value of evidence that were unsupported by empirical data. The Dotson case was an example of the first type. The analyst testified that Dotson was included in 11% of the population that could have been the semen donor, when in fact 100% of the population could have been the donor. An example of the second type of invalid testimony was in Timothy Durham's case, where the analyst opined that the particular reddish-yellow hue of his hair and the crime scene hair were found in "about 5 percent of the population."²¹ No empirical data exist on the frequency of hair characteristics, and thus that statement was totally unsupported.

As courts have long recognized, forensic expert testimony can play an important role in criminal trials. Juries may give special weight to testimony by forensic scientists; the Supreme Court has cautioned that "[e]xpert evidence can be both powerful and quite misleading because of the difficulty in evaluating it."²² These crimi-

²⁰ The findings regarding each transcript are summarized in the Appendix, which along with the transcripts themselves, has been made available online.

²¹ See *infra* note 157.

²² *Daubert*, 509 U.S. at 595; see also *United States v. Frazier*, 387 F.3d 1244, 1263 (11th Cir. 2004) ("[E]xpert testimony may be assigned talismanic significance in the eyes of lay jurors, and, therefore, the district courts must take care to weigh the value of such evidence against its potential to mislead or confuse."); *United States v. Hines*, 55 F. Supp. 2d 62, 64 (D. Mass. 1999) ("[A] certain patina attaches to an expert's tes-

nal trials all involved serious charges, typically rape and murder, and ten resulted in death sentences. This study makes no causal claims, however, regarding the degree to which invalid testimony contributed to wrongful convictions. Not only do we not know how jurors reached their verdicts, but these convictions were almost always supported by evidence in addition to the forensic evidence.

The advent of DNA technology has not solved the problem of invalid forensic testimony. DNA has replaced some, but not most, traditional forensic methods. Although DNA testing is now widely available in the kinds of sexual assault cases chiefly examined here, it is used in a small minority of criminal investigations. In a robbery, there is typically no semen deposited by the thief; in a drive-by shooting, no blood from the shooter may be left behind. In the overwhelming majority of cases, laboratories utilize additional forensic individualization disciplines other than DNA, some which are not unlike those that are the main subject of this study. Only two percent of law enforcement requests to crime labs involve requests for DNA analysis.²³ Nor is DNA analysis immune from inaccurate presentation of results. Several recent exonerations in our study set involved invalid trial testimony concerning DNA testing. Furthermore, this study describes only trial testimony. The incidence of faulty use or mischaracterization of the underlying data cannot be known without retesting or reexamination of the underlying forensic evidence.²⁴ Similarly, this study makes no conclusions about the state of mind of these analysts, which also cannot typically be known.

Unfortunately, our criminal system may not be well situated to prevent unscientific testimony. The adversarial system largely failed to police the invalid testimony during these trials. Defense counsel rarely cross-examined analysts concerning invalid testi-

timony unlike any other witness; this is 'science,' a professional's judgment, the jury may think, and give more credence to the testimony than it may deserve.'").

²³ See Joseph L. Peterson & Matthew J. Hickman, Bureau of Justice Statistics, *Census of Publicly Funded Forensic Crime Laboratories* 6 (2002), <http://www.ojp.usdoj.gov/bjs/pub/pdf/cpffcl02.pdf>; see also Cal. Comm. on the Fair Administration of Justice, *Report and Recommendations Regarding Forensic Science Evidence* 3 (2007) (referencing Barry Fisher, Director, Crime Laboratory for the Los Angeles County Sheriff Department).

²⁴ See *infra* Subsection II.F.2 (discussing few cases where retesting uncovered errors).

2009]

Invalid Forensic Science Testimony

11

mony and rarely retained experts, since courts routinely deny funding for defense experts. Prosecutors, moreover, presented erroneous accounts of the forensic evidence during closing arguments. In a few cases in which the defense challenged invalid forensic science, judges seldom provided relief. Courts do not typically review testimony after finding the underlying methodology reliable and permitting the forensic analyst to take the stand. As the NAS Report explained, “the legal system is ill-equipped to correct the problems of the forensic science community.”²⁵

For those reasons, the scientific community is a crucial source for both research and reform. Future research should examine the incidence of invalid testimony in cases beyond the cases examined, such as cases not involving DNA exonerations, cases involving more recent criminal trials, cases in which there is no DNA available to exonerate or confirm guilt, and cases involving different forensic disciplines. More important, the scientific community should respond in a forward-looking way by not just revisiting old cases, but also by issuing national standards for written reports and testimony in the future. Currently, no national or widely accepted set of standards for forensic science written reports or testimony exists. No entity promulgates such standards or ensures that all analysts adhere to standards for permissible scientific conclusions regarding forensic evidence. The NAS Committee report examining the needs of the forensic science community provides an important opportunity for legislators, lawyers, and scientists to implement such oversight mechanisms to ensure the accurate use of forensic science in the courtroom.

This Article will proceed as follows. Part I will summarize the findings and describe both the study method and background legal and ethical principles involved. Part II will present the findings by examining each type of invalid forensic science testimony and analysis, beginning with findings regarding conventional serology and microscopic hair comparison and proceeding to findings related to additional forensic science disciplines. Part III will describe the roles of defense counsel, prosecutors, and courts, and then conclude by recommending the adoption of national standards and

²⁵ Strengthening Forensic Science, *supra*, at 1-14; see also *id.*, ch.3.

oversight mechanisms to ensure that forensic science reports and testimony adhere to valid scientific standards.

I. SUMMARY OF FINDINGS, METHODOLOGY, AND PRINCIPLES

A. *The Study Set and Summary of Findings*

In 137 exonerees' trials—the group referred to below as the “study set”—trial transcripts were obtained in which forensic analysts were called to testify by the prosecution. The study set is a subset of the DNA exonerees as a whole. A total of 232 people have now been exonerated by post-conviction DNA testing.²⁶ One hundred fifty-six exonerees were identified as having trials in which forensic evidence was presented (three more pleaded guilty).²⁷ Efforts were made to obtain trial transcripts for the first 220 exonerees by contacting post-conviction attorneys, innocence projects and court clerks.²⁸ Of the 156 exonerees identified as having had fo-

²⁶ The study group only includes individuals who were exonerated, meaning that their conviction was vacated by a court or they received an executive pardon after DNA test results excluded them, and they were not retried. That group does not include non-DNA exonerations, including persons exonerated based on non-DNA forensic science. For additional discussion of the meaning of the term “exoneration,” see Garrett, *supra* note 7, at 64 n.33.

²⁷ This is a higher proportion of exonerations than previously reported. The first author's *Judging Innocence* study of the first 200 post-conviction DNA exonerations identified 113 cases supported by forensic science, or 57% of that sample set. See Garrett, *supra* note 7, at 81. There are several reasons why far more cases were identified in this study. First, cases were identified among the most recent exonerations.

Second, once trial transcripts were obtained and reviewed, new cases were identified that contained forensic science testimony. *Judging Innocence* did not examine such trial records, but rather judicial decisions and news reports. See *id.* at 66. Those sources did not mention that there was forensic science testimony during some of these trials.

Third, *Judging Innocence* examined only cases in which forensic evidence supported the state's case, because there the focus of the study was on whether that evidence was challenged post-trial—obviously an exoneree would not challenge exculpatory evidence. This study, because it focuses on the trial testimony itself, also includes 19 cases in which the state introduced forensic evidence at trial, even though that evidence did not support the state's case, but was rather non-probative or exculpatory. This study includes such cases to present a balanced picture of the testimony concerning forensic science. After all, many of the cases with invalid testimony should properly have been cases in which the forensic science was presented as non-probative or exculpatory.

²⁸ The authors stopped making systematic efforts to locate additional materials as this Article approached publication in October 2008. The authors note that one additional transcript not included in the study set has been obtained since that time: that

2009]

Invalid Forensic Science Testimony

13

rensic testimony at their trial, trial transcripts for 137 exonerees are studied here (14 were not located).²⁹ The 137 exonerees were convicted of the following crimes: rape (95 individuals), both rape and murder (33), murder (8), and attempted murder (1). Thus, the vast majority (128, or 93%) of the cases in the study set involved a sexual assault.

The testimony of forensic analysts in the 137 trials in the study set was reviewed, as summarized in the Appendix;³⁰ the transcripts have been made available online.³¹ In each of these 137 trials, forensic analysts were called to testify by the prosecution. Most of those analysts were employed by state or local law enforcement crime laboratories.³² There are over 350 crime laboratories in the United States.³³ The vast majority are operated by law enforcement

of David J. Bryson; that transcript included invalid testimony concerning hair comparison. At Bryson's trial, analyst Joyce Gilchrist testified that "it would be impossible not to be able to distinguish hairs from two different individuals," in effect asserting that human hairs are microscopically unique. Trial Transcript at 341, *State of Oklahoma v. David Johns Bryson*, No. CRF-82-5031 (Okla. D. Ct., Feb. 7, 1983).

The authors thank Michelle Morris and Kent Olson, reference librarians at the University of Virginia School of Law, for their extraordinary efforts to locate trial transcripts. The authors also thank Winston & Strawn, LLP for parallel efforts to locate each exoneree's trial transcripts.

²⁹ The 14 transcripts could not be obtained because no transcript was prepared on appeal, the transcript had been lost, or the case has been sealed. An additional 5 exonerees' trials involved only defense experts, whose testimony focused on exculpatory evidence. As discussed further, 19 trials had at least some defense expert testimony, and all 19 testified properly, often identifying flaws in testimony by analysts called by the state. See *infra* Subsection I.A.2.

³⁰ Throughout, this Article uses the term "forensic analysts" to refer generally to persons providing expert testimony regarding forensic evidence at trial. Those experts may be doctors, dentists, criminalists, police examiners, or have other professional titles.

³¹ Exoneree Trials: Testimony by Forensic Analysts, http://www.law.virginia.edu/html/librarysite/garrett_exoneree.htm. The Appendix, trial materials, and other reports associated with exonerees' cases cited in this paper can all be found at the webpage. The authors thank Jon Ashley and Mary Wood for their invaluable assistance in creating the webpage.

³² The exceptions were in cases in which FBI employees or analysts from private firms testified. For a history of the development of crime laboratories in the United States, see Paul C. Giannelli, *Regulating Crime Laboratories: The Impact Of DNA Evidence*, 15 *J.L. & Pol'y* 59, 61-67, 72 (2007).

³³ *Id.* at 70; see also Peterson & Hickman, *supra* note 23, at 2.

agencies as state or regional laboratories, though some are operated by local governments in large metropolitan areas.³⁴

In conducting a review of these 137 exonerees' trial transcripts, this study found invalid forensic science testimony was not just common but prevalent. This study found that 82 cases—60% of the 137 in the study set—involved invalid forensic science testimony.

This study focused on trial testimony, but noted instances in which it was later uncovered that the analyst withheld exculpatory forensic evidence. Withholding is not apparent from the trial testimony, but in 13 cases the concealment was later uncovered during post-conviction review, investigations, or civil discovery.³⁵ Three of those cases did not involve invalid testimony; thus a total of 85 cases—or 63% of the 137 cases—involved either invalid testimony or withholding of exculpatory evidence.

The testimony at these 137 exonerees' criminal trials chiefly involved serological analysis (100 cases) and microscopic hair comparison (65), because most of these cases involved sexual assaults for which such evidence was commonly available at the time. Indeed, in many cases, where both hair and semen were recovered from the crime scene, both disciplines were utilized. Some cases also involved testimony concerning: fingerprint comparison (13 cases), DNA analysis (11), forensic geology (soil comparison) (6), forensic odontology (bite mark comparison) (6), shoe print comparison (4), fiber comparison (2), voice comparison (1), and fingernail comparison (1).

In the two main categories of evidence present in the study set, serology and hair comparison testimony, this study found the following: Of the 100 cases involving serology in which transcripts were located, 57 cases, or 57%, had invalid forensic science testimony. Of the 65 cases involving microscopic hair comparison in

³⁴ There are, in addition, several federal laboratories, most notably the FBI lab, which is "the Nation's largest publicly funded forensic crime laboratory." Peterson & Hickman, *supra* note 23, at 2.

³⁵ For a discussion of these cases, see *infra* Section II.F. Thirteen total cases involved concealment of forensic evidence or analysis. Eleven also involved invalid testimony. These cases do not include at least 5 exonerees' cases in which it was withheld at trial that a prosecution witness had been hypnotized. Those cases are those of E. Honaker, L. Jean, L. Mayes, and G. Woodall. Forensic use of hypnosis involves uses unrelated to the identity of the perpetrator of a crime.

2009]

Invalid Forensic Science Testimony

15

which transcripts were located, 25 cases, or 38%, had invalid forensic science testimony.

Table 1, below, summarizes the incidence of invalid trial testimony by forensic analysts in the cases for which transcripts were located. (Ten cases involved more than one type of invalid testimony.)

Type of Forensic Analysis	Cases with trial transcripts	Cases involving invalid science testimony	Percentage of cases with trial transcripts involving invalid science testimony
Serology	100	57	57
Hair comparison	65	25	38
Soil comparison	6	0	0
Fingerprint comparison	13	1	8
Bite mark comparison	6	4	67
Shoe print comparison	3	1	33
DNA testing	11	3	27
Voice comparison	1	1	100

Table 1: Invalid Forensic Science Testimony by Type of Analysis

The cases involving proper testimony are also useful to examine. Many of those cases involved non-inculpatory testimony. Of the 55 cases in which all testimony was valid, 22 contained the testimony of forensic analysts who presented only evidence that was non-probative (13 cases) or exculpatory (11). Thus, almost half of the valid forensic testimony was not inculpatory and likely did not significantly support the conviction.

In contrast, most of the invalid forensic testimony involved evidence presented as inculpatory. In just 2 of the 82 cases with invalid testimony, the analysts testified that all of the forensic evidence was non-probative or inconclusive; in fact that evidence was exculpatory. The forensic testimony would have played a reduced role in many more of the 82 cases had forensic analysts accurately presented the evidence.

Three additional cases for which materials were obtained involved guilty pleas and no trial transcript. Two of those cases also involved invalid forensic analysis later exhibited in criminal trials

of co-defendants.³⁶ Those cases should trouble us since the vast majority of criminal cases are resolved through guilty pleas.³⁷

B. Study Protocol and Types of Invalid Testimony Identified

The authors established a protocol in advance to review the testimony, and created categories used to evaluate each transcript.³⁸ The authors were the primary reviewers of these transcripts,³⁹ but law student research assistants unfamiliar with these cases were trained on the protocol and reviewed each case as well.⁴⁰ As noted in the introduction, two basic categories of invalid science testimony recurred in these cases: (1) the misuse of empirical population data and (2) conclusions regarding the probative value of evidence in the absence of empirical data. The study protocol further divided testimony into three sub-types for each of those two categories. The Appendix lists how each case was categorized. The testimony itself is available for review online and Part II describes examples of each type of testimony. Below are the six types of invalid testimony that were identified.

1. Non-Probative Evidence Presented as Probative

The first category is the inaccurate use of empirical population data. The first and most common type of invalid testimony in this

³⁶ The cases of Christopher Ochoa and Bradford Marcellus each contained false confessions and involved invalid serology analysis later introduced in trials of co-defendants. The third case, that of James Ochoa, included DNA analysis and fingerprint analysis excluding him, but also dog scent identifications of him (although it is equivocal whether dog scent identification should be considered a form of forensic analysis).

³⁷ See discussion in Garrett, *supra* note 7, at 74 (“All but the nine who pleaded guilty in the innocence group (96%) were convicted at criminal trials. In contrast, 68% of murder convictions and 84% of felony rape convictions were obtained through plea bargaining.”).

³⁸ Edward Blake, a forensic scientist, and scientists including Eric Lander and Richard Lewontin reviewed these categories.

³⁹ Both authors have represented exonerees included in the study sample. When in law practice, Garrett assisted with civil cases brought by four of these exonerees. Neufeld and the Innocence Project that he co-directs assisted in the exonerations of many of these exonerees.

⁴⁰ The students did not conduct a review that was blind to the authors’ coding. However, they were instructed to review whether the transcripts were coded properly, and they reviewed the full testimony by each analyst.

category, present in 48 cases, was the interpretation of non-probative evidence as inculpatory evidence. That is, the testimony disregarded that the evidence was non-probative, and instead the analyst provided a statistic purporting to include the defendant and implying that a percentage of the population was excluded. The Dotson case described earlier provides an example of this type of invalid testimony. In a typical rape case, the evidence most likely to provide information about the genetic markers of the rapist is the vaginal swab collected immediately after the rape, when the victim is examined at a hospital. The analyst tested that evidence and testified that both Dotson and the semen donor possessed the B blood type, a type shared by only 11% of Caucasians. Eleven percent of Caucasians possess the B type; well-defined databases, developed over decades, provided the distribution of the four ABO blood group types in various racial and ethnic groups. However, that 11% statistic was invalid in the context of a rape prosecution, for it was not the combined frequency of all possible blood group types potentially possessed by the semen donor. Unlike today's DNA testing that can isolate and amplify very small amounts of genetic material, a major shortcoming of conventional blood grouping was that one could not separate the female contribution from the semen in a mixed stain present in a typical rape case. Therefore, one would generally not know if there was sufficient semen in the sample such that one would expect to detect its genetic markers. If there was not enough semen in the sample, only the victim's genetic markers would be observed. The analyst did not tell the jury that because the victim was also Type B, where her fluids were mixed in the sample, her Type B blood group substances could have masked any substances from the semen. The evidence was totally non-probative. In the Dotson case, the analyst should have told the jury that 100% of males could have been the donor. Part II describes this type of invalid testimony further; it involves the well-known problem of masking and non-quantification.⁴¹

In a related set of serology cases, moreover, the analysts testified that they observed no blood group substances in the crime scene samples. Rather than conclude that the contributor could have been any type because the evidence was potentially degraded,

⁴¹ See *infra* Subsection II.A.1.

these analysts testified that a defendant who did not secrete blood group substances was affirmatively included.

2. Exculpatory Evidence Discounted

A second type of invalid testimony occurred in 23 cases in which exculpatory evidence was discounted. For example, in Paul Kordonowy's case, serological tests of the victim's underpants revealed Type A antigens, which neither the victim nor Kordonowy possessed. Rather than testify that Kordonowy was excluded by the finding inconsistent with his type, the analyst told the jury to disregard that exculpatory evidence, and instead made the unsupported claim that bacteria could somehow have changed the reading and produced the Type A antigens.⁴² In other serology cases and several cases involving hair comparison, analysts similarly discounted exculpatory results and claimed to reach a non-probative or inculpatory result, often by relying on pure speculation.

3. Inaccurate Frequency or Statistic Presented

In a third type of invalid testimony present in 13 cases, the frequency or statistic presented was erroneous. In several exonerees' trials, analysts falsely divided frequencies in half. For example, in the Perry Mitchell case, the semen was left by a Type O secretor, and Type O secretors comprise 35% of the population. The serologist divided the accurate frequency in half and testified that only 17.5% of men could have contributed the semen and thus 82.5% of the relevant population was excluded.⁴³ However, population statistics regarding ABO blood group substances are identical for both sexes; 35% of both men and women are Type O secretors; thus, it was erroneous to divide that statistic in half.

4. Statistic Provided Without Empirical Support

The second major category of invalid testimony includes conclusions unsupported by any empirical data. In a fourth type of invalid testimony present in 5 cases, statements were made providing a frequency or probability in the absence of any empirical support.

⁴² See *infra* Subsection II.A.2.

⁴³ See *infra* Subsection II.A.3.

Some forensic disciplines involve more subjective analyses not premised on empirical population data. For example, in the field of microscopic hair comparison, no adequate empirical data exist regarding the frequency of microscopic characteristics of human hairs. Yet in the Bromgard case, the analyst testified that there was a 1 in 10,000 chance that the two hairs found at the crime scene could come from someone other than Bromgard.⁴⁴ Those frequency statistics were simply made up by the analyst.⁴⁵

5. Non-numerical Statements Provided Without Empirical Support

In a fifth type of invalid testimony present in 19 cases, non-numerical statements of probability or frequency were offered despite a lack of any empirical data. In the field of microscopic hair comparison, due to the lack of empirical data, the field adopted standards that the strongest statement of association that can be made by an analyst is that the hairs in question are “consistent” with the defendant’s or “could have” come from the defendant.⁴⁶ All analyst testimony, therefore, stating that a crime scene hair was “highly likely” to have come, “very probably” came, or did come from the defendant violates the basic scientific criterion that expressions of probability must be supported by data. For example, in the Calvin Scott case, the analyst testified that the chance that another person could have similar hair was remote, explaining, “I would not give a figure. It would be quite large.”⁴⁷ Use of such probability, frequency, or other individualizing statements was unsupported.

⁴⁴ See *infra* Subsection II.B.1.

⁴⁵ See *infra* Section II.B.

⁴⁶ To say that two items are “consistent” without being able to tell the jury that consistency is rare or common, renders the evidence potentially misleading and hence raises questions whether it is inadmissible as both irrelevant and unduly prejudicial. This study does not address evidentiary criteria—such as whether such testimony would be admissible under Federal Rules of Evidence 402 or 403—nor whether such testimony would satisfy *Daubert*. Other commentators have done so and courts should examine such questions carefully. See *infra* Section II.B. This study, however, is concerned only with the scientific validity of the testimony.

⁴⁷ See *infra* Subsection II.B.2.

6. Conclusion that Evidence Originated from Defendant

The sixth and final type of invalid testimony, present in 6 cases, claimed that the evidence did in fact come from the defendant and was unique to the defendant, despite no empirical data permitting such conclusions. For example, in Ray Krone's case, the analyst testified that the bite marks did in fact come from Krone's teeth, telling the jury, "that tooth caused that injury."⁴⁸ In two other cases the forensic odontologists (forensic dentists) were unequivocal that the defendants' teeth made the bite marks on the victim.⁴⁹ Forensic disciplines involving impression evidence, such as bite mark and shoe print comparison, have not developed any objective criteria at all by which to judge assertions about the likelihood that crime scene evidence came from a particular defendant.⁵⁰ Nor do any empirical data exist to support a claim that a bite mark is uniquely identifiable as belonging to a particular person.

* * *

These six types of invalid testimony may occur in other disciplines not reviewed here, and conversely, additional types of invalid forensic testimony may occur in cases not in the study set. As noted at the outset, this study cannot speak to questions concerning how often invalid forensic science testimony occurs in other types of more typical criminal cases. The study set is limited not only to DNA exonerees, but also to trials resulting chiefly in rape or rape and murder convictions in the 1980s. Perhaps such cases involving felonies in contentious cases that proceeded to trial were more likely to involve pressures on the state to overstate the evidence, including forensic evidence, making these cases unrepresentative of more common and less serious crimes. On the other hand, perhaps such cases did not involve such pressure to overstate forensic evidence. Perhaps there would be little pressure to overstate forensic evidence if the defense did not meaningfully contest forensic evidence. If so, these cases might be representative of a more

⁴⁸ See *infra* Section II.D.

⁴⁹ See *infra* Section II.D.

⁵⁰ See Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science*, 61 *Vand. L. Rev.* 199 (2008).

widespread problem. Those questions can not be answered by reviewing just the trials of DNA exonerees.

After all, the particular forms of forensic analysis reviewed reflect the make-up of the cases in the study set. Almost all exonerees in the study set were convicted of rape or rape and murder. This is not because the quality of forensic testimony is worse in rape cases, but rather because DNA testing could be later used to identify the source of the semen left by a rapist, which is usually dispositive of guilt in cases involving stranger-perpetrators in which the central issue is the identity of the assailant. Most of the exonerees were convicted of crimes involving strangers in which the identity of the perpetrator was in question at trial, and ultimately it was shown through post-conviction DNA testing that in fact the wrong person was convicted. Therefore, the study set disproportionately included evidence that one would expect in a rape case: serology analysis of material collected as part of a rape kit and microscopic hair comparison of hairs found at the crime scene, often from combings of the victim or the victim's clothes.

This explains why so many of the trials studied involved semen or hair evidence and also why there were few trials studied involving fiber analysis, bite marks, fingerprints, toolmark analysis, and other forensic disciplines. Such other forensic disciplines do not routinely examine evidence common in a sexual assault case.⁵¹ Nor does evidence such as a latent fingerprint typically have preserved relevant biological evidence that can later be tested using DNA analysis.⁵² These exonerees all had cases in which such evidence

⁵¹ The one study to examine which types of forensic testing were conducted in different types of felony investigations supports this conclusion. See Joseph L. Peterson et al., *Forensic Evidence and the Police, 1976–1980*, Nat'l Archive of Crim. Just. Data, Inter-University Consortium for Political and Social Research, Study No. 8186 (1985). That study developed types of forensic analysis in over 1700 felony investigations conducted in four urban police departments from 1976 through 1980. The data sheets were analyzed with the always outstanding assistance of University of Virginia Reference Librarian Jon Ashley. The 183 rape cases in that set chiefly had serological analysis of blood (68 cases), semen (153), or hair evidence (87 with pubic hair and 55 with head hair). *Id.* Comparatively few rape cases had fiber analysis (5 cases), latent print analysis (46), bullet analysis (4), or impression analysis (5). *Id.* In contrast, as one would expect, few of the 223 murder cases had semen analysis (5 cases), but many involved analysis of bullets (142) or latent fingerprints (94).

⁵² In the Peterson data set, the cases with biological evidence, such as serology evidence, were far more common in cases with hair evidence (86%) than in cases with,

was in fact collected at the crime scene, because each was exonerated when the DNA testing was later conducted on that material, typically from a rape kit or on certain hair evidence.⁵³ This study also does not include forensic analysis unrelated to the issue of identity introduced to show how a crime occurred or that it occurred, such as autopsy evidence.⁵⁴ Thus, the role of particular forensic disciplines as well as the role of invalid forensic science in the cases studied here would be different for other types of criminal cases, and even for other types of cases in which identity is at issue. For example, the study set cases did not typically involve analysis of bullets that one would expect in cases involving shootings, or tire tread analysis that one would expect in cases involving vehicular assault.

Forensic evidence in the vast majority of criminal cases that result in guilty pleas does not receive the scrutiny of a trial. However, the set of DNA exonerees in this study consists of persons convicted at a trial. The cases studied here not only involved trials, but they mostly involved trials in the 1980s. Today, issues of identity in sexual assault cases may often be resolved through DNA testing pre-trial, making it less likely that some of the invalid testimony observed regarding hair comparison or serology would occur.⁵⁵ However, other non-sexual assault cases are not as susceptible to DNA testing and may present some of the same issues implicated

for example, fingerprint evidence (28%). *Id.* Further, “although it is now possible, in the laboratory, to extract DNA from a fingerprint, this has not been done in the field, and it would certainly not be possible with a fingerprint that has aged in an evidence locker.” Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 *J. Crim. L. & Criminology* 985, 1026 (2005) (citation omitted).

⁵³ See Garrett, *supra* note 7, at 73 (“The 200 exonerees were charged and convicted chiefly of rape (71%), murder (6%), or both murder and rape (22%). This is not surprising; rape cases in particular often have relevant biological material for DNA testing.”).

⁵⁴ False pathology evidence, for example, could lead to wrongful convictions where no murder in fact occurred but rather the death was due to natural causes. See, e.g., Mark Bonokoski, Editorial, *Experts Must be Impartial*, *The Daily Observer*, Feb. 7, 2008 (describing the work of Charles Smith, “an expert witness (supposedly) in forensic pathology who lied, invented, forgot, pretended, withheld, dismissed, neglected, guessed—and, as a result, sent many people to jail for crimes that never happened”). Cases in which no crime in fact occurred do not raise issues regarding the identity of the perpetrator for which post-conviction DNA testing would lead to exoneration. As a result, no such cases were present in the study set.

⁵⁵ See Brandon L. Garrett, *Claiming Innocence*, 92 *Minn. L. Rev.* 1629, 1634 (2008).

in these exonerees' cases. Latent fingerprint comparison, for example, is still in wide use; indeed it is used far more often than DNA testing.⁵⁶ New forensic techniques continue to be developed that involve the same sorts of subjective comparison not grounded in empirical data, which might then risk invalid testimony if analysts do not inform the jury that no probability or frequency can be supported. For example, the FBI is developing a new technology which it claims can identify unique characteristics of human voices.⁵⁷

Further study is necessary to assess questions regarding incidence of invalid testimony in recent trials, in trials not involving sexual assaults, and in trials not involving wrongful convictions. The next Section describes what limited information is available concerning such questions.

C. Questioning the Incidence of Invalid Forensic Testimony

Senator Orrin Hatch, commenting on the need to provide new resources for forensic sciences, referring to the fraudulent work of Oklahoma City police department forensic analyst Joyce Gilchrist that contributed to several wrongful convictions, noted:

[W]e are all troubled by allegations that mistakes by a police chemist in Oklahoma helped send innocent people to prison. This isolated situation should not be used unfairly to indict the thousands of forensic scientists who perform their work professionally and responsibly. It should, however, remind us that those who work in our criminal justice system have an obligation to be diligent, honest, and fair-minded.⁵⁸

While not disagreeing with that statement, this study describes how the invalid testimony in DNA exoneration cases did not just

⁵⁶ See Integrated Automated Fingerprint Identification System, <http://www.fbi.gov/hq/cjisd/iafis.htm> [hereinafter IAFIS]; Peterson & Hickman, *supra* note 23, at 6–9.

⁵⁷ Dina Temple-Raston, Voice “Fingerprints” Change Crime-Solving, Nat’l Pub. Radio, Jan 18, 2008, <http://www.npr.org/templates/story/story.php?storyId=18479095>.

⁵⁸ See DNA Crime Labs: The Paul Coverdell National Forensic Sciences Improvement Act: Hearing Before the Comm. on the Judiciary, 107th Cong. 2–3 (2001) (statement of Sen. Orrin G. Hatch, Chairman, Comm. on the Judiciary), available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_senate_hearings&docid=f:78008.pdf.

involve a few “bad apples,” like Gilchrist, who have been the subject of high profile investigations.⁵⁹ Several forensic analysts testified in more than one trial in the study set, including Pamela Fish (5 trials), Arnold Melnikoff (3), Joyce Gilchrist (3), and Fred Zain (6). However, 61 of the analysts who delivered invalid testimony did so in just one trial in the study set. The study set included invalid testimony by 72 forensic analysts called by the prosecution and employed by 52 laboratories, practices, or hospitals from 25 states.⁶⁰

This study does not examine the state of mind of forensic analysts.⁶¹ Invalid testimony could be explained not by intentional or reckless acts, but rather by inexperience, poor training, or inadequate supervision.⁶² If these particular analysts lacked adequate training or supervision, then one wonders about their testimony in other cases as well as testimony by their colleagues. Most crime laboratories do not employ more than a dozen analysts; each one of these analysts could have testified in many cases each year.⁶³ Indeed, in many of the trials studied, the analysts, when describing their credentials, stated that they had testified on numerous occasions, sometimes even in hundreds of trials.⁶⁴

⁵⁹ See William C. Thompson, *Beyond Bad Apples: Analyzing the Role of Forensic Science in Wrongful Convictions*, 37 Sw. U. L. Rev. 101, 112-119 (2009).

⁶⁰ The states are: Arizona, California, Connecticut, Florida, Georgia, Idaho, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Mississippi, Missouri, Montana, Nevada, New Jersey, New York, Ohio, Oklahoma, Pennsylvania, South Carolina, Texas, Virginia, and West Virginia.

⁶¹ Other studies develop questions regarding possible bias or observer effects, where forensic analysts are typically employed by law enforcement. D. Michael Risinger et al., *The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 Cal. L. Rev. 1, 5-6 (2002).

⁶² Samuel R. Gross, *Expert Evidence*, 1991 Wis. L. Rev. 1113, 1178 (“The worst that can be said about an expert opinion is not that it is a lie—that criticism is often beside the point—but that it is unreasonable, that no competent expert in the field would hold it. Correspondingly, the most dangerous expert witness is not one who lies (although she may do that too), but one who is ignorant or irresponsible.”).

⁶³ See Peterson & Hickman, *supra* note 23, at 3.

⁶⁴ For example, David Brody, in Neil Miller’s case, testified that he had testified in court during his career “at least a thousand times.” Trial Transcript at 190, *Commonwealth v. Neil Miller*, No. 92-P-612 (Okla. Dist. Ct. Dec. 14, 1990) [hereinafter Transcript, N. Miller]. Elmer Gist testified in the Honaker case that he had “testified in western Virginia, northern Virginia, the Tidewater area, approximately a hundred and eighty times total.” Trial Transcript at 212, *Commonwealth v. Edward William Honaker*, No. 2159-75 (Va. Cir. Ct., April 10, 1985) [hereinafter Transcript, Honaker].

Nor does the prevalence of invalid forensic testimony in these trials speak to what “caused” these wrongful convictions. Though each case involved an erroneous outcome—an innocent person convicted—invalid forensic science testimony may not have been the deciding factor leading juries to convict. In addition to the forensic evidence, other evidence—particularly eyewitness identifications—supported most of these convictions. Forensic science testimony might not by itself “cause” a conviction where criminal trials typically involve multiple pieces of evidence and actions by several actors.⁶⁵ For example, Gary Dotson might still have been convicted even if the forensic analyst had correctly observed that any male could have been the semen donor. Among other evidence in the case, the victim had identified him as the rapist. The forensic analyst’s invalid forensic testimony did serve some role in buttressing the false eyewitness identification, yet one cannot typically know how jurors weighed the evidence in reaching the decision to convict. As noted, several of these trials involved forensic evidence—in a few cases DNA evidence—that excluded the defendant, and yet the state still secured the conviction. However, courts and scholars have long recognized that jurors may place special trust in scientific evidence.⁶⁶ Studies also suggest that the manner in which the forensic evidence is presented to the jury impacts how jurors weigh that evidence.⁶⁷

⁶⁵ See Garrett, *supra* note 7, § II.A.

⁶⁶ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 595 (1993); see also *United States v. Frazier*, 387 F.3d 1244, 1263 (11th Cir. 2004) (“[E]xpert testimony may be assigned talismanic significance in the eyes of lay jurors, and, therefore, the district courts must take care to weigh the value of such evidence against its potential to mislead or confuse.”); *United States v. Hines*, 55 F. Supp. 2d 62, 64 (D. Mass. 1999) (“[A] certain patina attaches to an expert’s testimony unlike any other witness; this is ‘science,’ a professional’s judgment, the jury may think, and give more credence to the testimony than it may deserve.”); Gross, *supra* note 62, at 1179–81 (reviewing empirical research regarding degree to which juries rely upon and comprehend expert evidence).

⁶⁷ See, e.g., Jonathan J. Koehler, *The Psychology of Numbers in The Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient*, 74 S. Cal. L. Rev. 1275 (2001); Dawn McQuiston-Surrett & Michael J. Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 *Hastings L.J.* 1159 (2008); John Monahan et al., *Communicating Violence Risk: Frequency Formats, Vivid Outcomes, and Forensic Settings*, 1 *Int’l J. Forensic Mental Health* 121, 126 (2002).

Two of the exonerees' cases involved bench trials, providing information about how the fact-finder reasoned. In the Willie Davidson case, invalid forensic science appeared to have misled the fact-finder. The judge emphasized that the guilty verdict was "supported by that scientific evidence and [the victim's] identification."⁶⁸ But when explaining the scientific evidence, he appeared confused, understandably so, because improper testimony concerning the serology had ignored the problem of masking and quantification. The judge stated: "Then it had the type of a non-secretor. The defendant is a non-secretor. That by itself isn't totally conclusive. Forty-two percent are of that, so that doesn't nail it down."⁶⁹ Actually, 42% was not a proper statistic. No male could be excluded by the serological techniques used at the time. Where the victim was Type O and Type O material was observed, the blood group substances could have solely originated from the victim, and thus any person could have been the semen donor. Separately, in Nathaniel Hatchett's case, powerful exculpatory forensic evidence was disregarded. DNA testing conducted before trial on the semen evidence from a single-perpetrator rape had excluded Hatchett. Nevertheless, the judge in the bench ruling found the DNA results not dispositive where Hatchett had confessed, stating, "in light of the overwhelming evidence that the Court has . . . the Court does not find that the laboratory analysis is a fact which would lead to a verdict of acquittal."⁷⁰

Again, this study's data do not support claims about the incidence of invalid forensic science testimony in cases outside of the 137 trials studied, but rather points to the need to investigate the nature of the problem. Some evidence from cases outside this study set also suggests that this problem deserves further attention, and that invalid forensic testimony may not be associated with wrongful convictions, but rather may be part of a different and larger problem. Studies have found high error rates in a series of forensic dis-

⁶⁸ Trial Transcript, *Commonwealth of Virginia v. Willie Davidson*, No. 919-81 (Va. Cir. Ct., May 27, 1981) (page numbers illegible).

⁶⁹ *Id.*

⁷⁰ Trial Transcript at 276-77, *State of Michigan v. Nathaniel Maurice Hatchett*, No. 97-1496-FC (Mich. Cir. Ct., Mar. 3, 1998).

ciplines.⁷¹ Such studies may shed light on the reliability of the underlying method or its application by forensic practitioners, but they do not shed light on whether trial testimony comports with scientific standards. Indeed, few have studied testimony by forensic analysts. One of the purposes of this Article is to encourage future efforts to review and improve the quality of forensic science testimony.

One reason that compilations of more systemic data concerning the quality of forensic testimony during criminal trials are lacking is that crime laboratories do not routinely collect or review such testimony. Even after these DNA exonerations, not only have investigations into these individual cases often not occurred, but investigations regarding systemic problems in laboratories remain rare. When our system has investigated laboratories in response to these exonerations, systemic problems have been uncovered. Noteworthy examples include the Houston Police Department investigation led by Michael Bromwich that uncovered hundreds of cases involving invalid serology analysis beyond the two post-conviction DNA exonerations that sparked the investigation.⁷² Similar audits have occurred in reaction to DNA exonerations at laboratories in Cleveland, Ohio, and Baltimore, Maryland, and laboratories in Oklahoma, West Virginia, and Virginia.⁷³

What little information does exist regarding cases outside our study sample does not provide cause for optimism. Simon Cole has conducted a preliminary effort, examining 34 transcripts involving latent print testimony, finding “over-claiming,” or expert testimony exaggerating its own probative value, prevalent in that group of cases.⁷⁴ Another example is the recent National Research Council report, which uncovered invalid testimony by FBI analysts who tes-

⁷¹ See Joseph L. Peterson & Penelope N. Markham, *Crime Laboratory Proficiency Testing Results, 1978–1991, II: Resolving Questions of Common Origin*, 40 *J. Forensic Sci.* 1009, 1010 (1995); see also *Evidence of Injustice*, CBS News, Sept. 14, 2008, at http://www.cbsnews.com/stories/2007/11/16/60minutes/main3512453_page4.shtml (noting that a review of 100 FBI cases with trial testimony concerning bullet lead analysis uncovered that “almost half” involved flawed testimony).

⁷² See Michael R. Bromwich, Executive Summary, *in* Fifth Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room 1–2 (May 11, 2006), available at <http://www.hpdlabinvestigation.org>.

⁷³ See Possley, *supra* note 9, § 1, at 1, 20–21.

⁷⁴ See Cole, *supra* note 16; see also Saks & Koehler, *supra* note 50, at 205–06.

tified for decades that bullets “must have come from the same box” without any empirical support.⁷⁵ The Detroit Police Department Crime Laboratory was recently closed based on a “systemic” failure to properly conduct firearms analysis, a type of analysis not studied here.⁷⁶ In several disciplines involving impression evidence, as developed below, the relevant disciplines provide guidelines regarding trial testimony that explicitly permit invalid testimony not based on empirical evidence.⁷⁷

Nor is it difficult to find a host of reported appellate decisions describing invalid forensic science testimony similar to that in these exonerees’ trials. Reported decisions regarding invalid serology, hair comparison, fingerprint comparison, and bite mark comparison testimony can readily be found on Westlaw, and numerous such cases are collected in treatises on scientific evidence.⁷⁸

Our quite preliminary effort to test whether the testimony in these exonerees’ trials is representative of testimony in similar trials suggests that invalid testimony was also common in trials in which there has been no DNA exoneration, involving similar rape and murder charges and from the same time period. To date, 30 trial transcripts in such “matched” cases have been collected from Missouri (10 transcripts), Texas (11), and Virginia (9). Almost two-thirds of those trials exhibited invalid forensic science testimony, including the same types observed in the exonerees’ trials, and including testimony by some of the same analysts who testified in the

⁷⁵ Comm. on Scientific Assessment of Bullet Lead Elemental Composition Comparison, Nat’l Research Council, *Forensic Analysis: Weighing Bullet Lead Evidence* 90–94 (2004); see also Cole, *supra* note 16, at 820.

⁷⁶ See Nick Bunkley, *Detroit Police Lab Is Closed After Audit Finds Serious Errors in Many Cases*, N.Y. Times, Sept. 25, 2008, at A17.

⁷⁷ See Michael Bowers, *The Scientific Status of Bitemark Comparisons*, in *Modern Scientific Evidence* 538, 549–50 (David L. Faigman et al. eds., 2002); Iain A. Pretty, *Reliability of Bitemark Evidence*, in *Bitemark Evidence* 531, 543 (Robert B.J. Dorion ed., 2005); see also Cole, *supra* note 16, at 820–22 (discussing “institutionalized overclaiming”).

⁷⁸ See *infra* notes 301–02; 2 Paul C. Giannelli & Edward L. Imwinkelried, *Scientific Evidence* § 24-3 (4th ed. 2007) (describing and citing to a “massive body of case law” admitting testimony regarding hair comparison, including testimony found here to be invalid, such as use of probabilistic statements); see also *id.* § 24-5 (describing reported cases reviewing fiber comparison testimony); 1 Paul C. Giannelli & Edward L. Imwinkelried, *Scientific Evidence* § 13-5 (4th ed. 2007) (digesting case law concerning bite mark comparison).

exonerees' trials.⁷⁹ Such matched cases likely do not involve innocent convicts, but rather guilty convicts who also had invalid forensic testimony presented at their trials.

Neither matched cases involving likely correct outcomes, nor most cases involving wrongful convictions, tell us about false negatives: cases in which invalid forensic analysis led to guilty persons going free. Studies of proficiency testing of forensic laboratories, however, suggest that false negatives are far more common than false positives, and also that error rates may be generally high across a wide range of forensic techniques, including those studied here.⁸⁰

Thus, preliminary evidence suggests that even if most of the forensic science testimony in DNA exonerees' trials was invalid, such invalid testimony may not be associated with wrongful convictions. More troubling, it may be a phenomenon in serious criminal trials generally, at least during the time period in question. However, that question can not be definitively answered nor can the more difficult question of whether such testimony is common in more typical criminal cases. Future research should investigate the incidence of invalid forensic science testimony.

⁷⁹ See Garrett, *supra* note 7, § I.B (explaining the method for selecting such "matched" cases with similar characteristics to those of the exonerees). In short, a "matched" case involved the same type of conviction in the same state and with reported decisions in the same years, but in which no DNA testing was later conducted to exonerate the defendant. The transcripts collected from these states in non-exoneration cases have been made available online at the same webpage at which the exoneree materials have been posted. Twenty of the cases involved serology testimony, 10 hair comparison, 5 fingerprint comparison, 2 bite mark comparison, and 3 involved testimony concerning DNA testing. Nineteen cases involved invalid testimony and one more involved concealment of exculpatory information that was uncovered post-trial. Thus, 63% involved invalid forensic science testimony, approximately the same percentage as among the trials of exonerees who had forensic science testimony at trial. Special thanks to Kent Olson and the Texas Defender Service for their invaluable assistance in locating these materials.

⁸⁰ Peterson & Markham, *supra* note 71, at 1009–11 (summarizing study results finding a series of forensic disciplines with better than 10% correct identifications in proficiency tests, but other disciplines with error rates in the 10–20% range or even higher error rates).

D. Ethics and Forensic Science Testimony

Forensic science is uniquely concerned with the introduction of evidence in courtrooms, particularly in criminal courts where the stakes can be extremely high. Thus, “criminalistics . . . has as its primary objective a determination of physical facts which may be significant in legal cases.”⁸¹ An ethical forensic analyst has a professional obligation not to mislead the jury during testimony at trial and not to mislead the state and defense when preparing forensic reports.

To the extent that a prosecutor or defense attorney asks questions that are misleading or confusing, “[t]he expert witness’s obligation . . . is to give a full and complete presentation of the opinion and the reasons for that opinion,” Peter Barnett writes, adding that “[t]actics on the part of either the witness or the lawyer that tend to obscure the testimony, limit the full disclosure of the basis for the testimony, or confuse or obscure the implications of the testimony are inappropriate and, under some circumstances, may be unethical or illegal.”⁸²

While no single ethical code applies to all practicing criminalists, much less all forensic analysts in the United States, a series of professional entities have promulgated ethical codes that shed light on testimony discussed here, including the American Board of Criminalists (“ABC”), the American Academy of Forensic Sciences (“AAFS”), and the California Association of Criminalists (“CAC”).⁸³ As a general matter, these codes counsel independent evaluation of the evidence and truthful and non-misleading testimony in court. The ABC Code of Ethics asks that all analysts ensure that opinions are rendered “only to the extent justified” by the evidence, and to ensure that their testimony is presented “in a clear, straightforward manner” that does not “extend themselves beyond their field of competence, phrasing their testimony in such a manner so that the results are not misinterpreted.”⁸⁴

⁸¹ The Code of Ethics of the California Association of Criminalists Preamble (1957), reprinted in Peter D. Barnett, *Ethics in Forensic Science: Professional Standards for the Practice of Criminalistics* 125 (2001).

⁸² *Id.* at 81.

⁸³ See *id.* at 7, 81.

⁸⁴ Am. Bd. of Criminalistics, Code of Ethics §§ 9–10, reprinted in Barnett, *supra* note 81, at 153. The Code also states that criminalists shall “[m]aintain an attitude of inde-

The AAFS Code simply forbids a “material misrepresentation of data upon which an expert opinion or conclusion is based.”⁸⁵ The AAFS Guidelines also adopt “good forensic practice guidelines,” which add that “[u]nlike attorneys, forensic scientists are not adversaries. They take an oath in court to tell the whole truth. They should make every effort to uphold that oath.”⁸⁶ Further, when presenting their opinions, “[e]very reasonable effort should be made to ensure that others (including attorneys) do not distort the forensic scientist’s opinions.”⁸⁷

The CAC Code does not apply to most of the analysts in this study set, but in contrast to the ABC and AAFS codes, it imposes far more rigorous requirements. The CAC Code states that “[i]n all respects, the criminalist will avoid the use of terms and opinions which will be assigned greater weight than are due them. Where an opinion requires qualification or explanation, it is not only proper but incumbent upon the witness to offer such qualification.”⁸⁸ The CAC Code requires that the expert indicate when an opinion “may lack the certainty of other opinions he might offer,” and will “leave no false impressions in the minds of the jurors.”⁸⁹ The CAC Code adds that an expert “will not . . . assign greater significance to an interpretation than is justified by the available data.”⁹⁰

pendence and impartiality in order to ensure an unbiased analysis of the evidence.”
Id. § 14.

⁸⁵ Am. Acad. of Forensic Sci., Code of Ethics and Conduct, § 3, *reprinted in* Barnett, *supra* note 81, at 123.

⁸⁶ Am. Acad. of Forensic Sci., Good Forensic Practice Guidelines § 13, *reprinted in* Barnett, *supra* note 81, at 144.

⁸⁷ Id. § 9.

⁸⁸ Cal. Ass’n of Criminalists, Code of Ethics § III.E, *reprinted in* Barnett, *supra* note 81, at 128.

⁸⁹ Id. § III.D.

⁹⁰ Id. § III.C. This study does not examine cases in which analysts made a “material misrepresentation of education, training, experience, or area of expertise.” Am. Acad. of Forensic Sci., *supra* note 85, § 2. There is evidence suggesting that this would be a useful area for future study. For example, Fred Zain, who testified in five cases in this study, had performed poorly in the basic FBI serology course in 1977. However, this was not included in his personnel file, and he was promoted to supervisor of the West Virginia State Police Crime Laboratory, Serology Division, shortly thereafter. See *In re Renewed Investigation of the State Police Crime Lab., Serology Div.*, 438 S.E.2d 501, 514–20 (W. Va. 1993) (noting also that “Zain may have testified falsely concerning his academic credentials”).

Those ethical rules do not provide guidance on the permissible scope of testimony within a particular discipline; they speak to the general norms of expert conduct. Thus, those rules do not provide any scientific standards governing courtroom testimony which are the focus of this study.

E. Legal Regulation of Forensic Testimony

Courts do not typically review the presentation of forensic science testimony during criminal trials. As noted, courts recognize that jurors place special trust in expert witnesses to explain applicable scientific principles. Courts therefore regulate the matters upon which experts may testify. Thus, while a police officer could identify a defendant as the person seen committing a crime, a forensic analyst may only testify regarding an identification of a defendant using forensic methods supported by sound science. The wrongful convictions in this study occurred chiefly in the 1980s, prior to the trilogy of Supreme Court decisions heightening reliability requirements for scientific and expert testimony.⁹¹ Under the *Frye v. United States* test that governed in federal courts and most states at the time of these convictions (since replaced in most jurisdictions by the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals*), courts would permit expert testimony based only on a methodology that was "generally accepted" as reliable in the relevant scientific community."⁹² Scholars have criticized "the stunning failure of judges to provide any sort of check" on unsupported forensic evidence, describing a failure to rigorously adhere to *Daubert's* standards in criminal cases.⁹³ This study does

⁹¹ *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999); *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997); *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

⁹² *Daubert*, 509 U.S. at 584 (quoting *Frye v. United States*, 54 App. D.C. 46, 47 (1923)); Alice B. Lustre, Annotation, Post-Daubert Standards for Admissibility of Scientific and Other Expert Evidence in State Courts, 90 A.L.R.5th 453, 481 (2001) (describing that the *Frye* approach is now a minority approach).

⁹³ David L. Faigman, Anecdotal Forensics, Phrenology, and Other Abject Lessons From the History of Science, 59 *Hastings L.J.* 979, 991-92 (2008). An analysis of the application of *Daubert* in its first decade reveals that while it was used frequently to exclude questionable scientific evidence in civil cases, it almost never resulted in the exclusion of forensic evidence proffered by the prosecution in criminal cases. Peter J. Neufeld, The (Near) Irrelevance of *Daubert* to Criminal Justice and Some Suggestions for Reform, 95 *Am. J. Pub. Health* S107, S109 (2005); see also D. Michael Ris-

not address that debate, because as those scholars point out, *Daubert* has not been carefully applied to regulate the subject of this study, the trial testimony of forensic analysts.

At least in criminal cases, having found that the underlying discipline is satisfactory and the evidence admissible following the *Frye*—or now the *Daubert*—standard, courts do not typically examine conclusions experts reach on the stand regarding whether statistical claims or other inferences drawn from the data are supported by the evidence.⁹⁴ There is no screening of the case specific inferences and opinions before the jury hears them. Yet it is precisely while the expert testifies that, as Simon Cole puts it, “the rubber meets the road,” and the jury hears claims about the actual evidence in the case.⁹⁵ In the few cases where the exonerees’ defense counsel raised objections to invalid forensic testimony, judges rarely limited it. When appellate attorneys challenged faulty forensic testimony, courts rarely granted relief, often finding any error to be harmless.⁹⁶

Thus, if an expert overstates the evidence or presents it in a misleading fashion, cross-examination is relied upon to test the evidence. Yet in a criminal case, the defense is typically an unarmed adversary that lacks expert assistance. Also of crucial importance, the presentation of forensic science during criminal trials is usually one-sided, provided only by analysts testifying for the prosecution. Most states do not routinely fund the provision of forensic experts for indigent defendants, though there are strong arguments that under *Ake v. Oklahoma* defendants should be entitled to expert assistance as a matter of due process, at least in some types of cases.⁹⁷ As a result, courts routinely deny indigent defendants the funds to

inger, Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?, 64 Alb. L. Rev. 99, 149 (2000).

⁹⁴ See Cole, supra note 16, at 819 (“[J]udges assume that their work is done once they have ruled proffered evidence admissible or inadmissible.”).

⁹⁵ Id. at 818.

⁹⁶ See infra Subsection III.A.3.

⁹⁷ See *Ake v. Oklahoma*, 470 U.S. 68, 83 (1985); 1 Giannelli & Imwinkelried, supra note 78, § 4-5; Paul C. Giannelli, *Ake v. Oklahoma: The Right to Expert Assistance in a Post-Daubert, Post-DNA World*, 89 Cornell L. Rev. 1305, 1338–41 (2004); Gross & Mnookin, supra note 18, at 189 (“In many criminal cases, there is only one side on expert issues: the prosecution. The result is a national scandal. We have seen case after case of systematic fraud and incompetence by prosecution experts and police crime laboratories, with no end in sight.”).

hire their own forensic experts. Almost all of the analysts testifying in the 137 exonerees' trials worked for police laboratories; only 19 exonerees retained experts.⁹⁸ The fact-finders in most of these cases were jurors: non-experts who could be easily misled by invalid scientific testimony. Prosecutors not only elicited invalid forensic testimony, but sometimes further misrepresented the forensic science in their closing arguments, perhaps leading the jury to draw incorrect conclusions in cases where the analyst provided proper testimony.⁹⁹

In addition to *Daubert*, a second legal rule applicable to state experts, the Supreme Court's decision in *Brady v. Maryland*, holds that the State violates the due process rights of a defendant by withholding material, exculpatory information from the defense.¹⁰⁰ Expert fabrication of evidence violates the Due Process Clause as well. For example, the Court unanimously held in *Miller v. Pate* that a conviction should be set aside where the State obtained a conviction based on testimony that certain stains on underwear owned by the defendant matched the victim's blood type but where it was later shown that the stains were paint.¹⁰¹ By its nature, concealed evidence rarely comes to light and violations are rarely detected, much less remedied.

Where courts do not regulate the content of expert testimony, and defendants typically do not have experts with which to effectively counter State-preferred forensic testimony in criminal trials, the scientific standards within the forensic sciences are the most important source for regulating the content of forensic science testimony. This Article next develops a series of examples in which analysts did not adhere to valid scientific standards. The Article concludes that existing regulations are not adequate to prevent invalid forensic science testimony.

II. RESULTS: INVALID FORENSIC SCIENCE TESTIMONY

The cases of the exonerees whose trials had forensic science testimony chiefly involved serology analysis of material collected as

⁹⁸ See *infra* Subsection III.A.2.

⁹⁹ See *infra* Subsection III.A.1.

¹⁰⁰ 373 U.S. 83, 87 (1963).

¹⁰¹ 386 U.S. 1, 5-7 (1967).

part of a rape kit and microscopic hair comparison of hairs found at the crime scene, often from combings of the victim or the victim's clothes. The Sections that follow first develop the use of serology and hair comparison in these exonerees' trials. Next, the Article discusses additional forensic disciplines employed in smaller numbers of these cases—namely, bite mark comparison, DNA testing, and fingerprint comparison. For each type of analysis, Sections below describe the types of invalid testimony present with illustrative examples of each.

A. Invalid Forensic Serology Testimony

Of the 137 trial transcripts in the study set, 100 had testimony regarding serology analysis. Of those, 57 involved invalid testimony, 46 of which involved “masking” and quantification problems, which will be described further below.

In the “serology era” prior to the advent of DNA testing technology, the most precise method for including or excluding an individual as the source of the biological evidence at a crime scene was conventional serology, which involves analysis of fluids for certain markers that are lifelong individual characteristics, chiefly based on water-soluble ABO blood group substances and the phosphoglucomutase (“PGM”) enzyme genetic marker system. The ABO blood group substances are found on the surface of red blood cells. In addition, water-soluble ABO blood group substances are expressed by about 80% of the population in other body fluids, including saliva, semen, and vaginal fluid;¹⁰² these individuals are called secretors.¹⁰³ Secretor status is a genetically determined trait. Analysts test fluids for the presence of the A, B, and H blood group substances using ABO typing, the first means developed for distinguishing individuals based on characteristics of their body fluids.¹⁰⁴ This conventional serology analysis cannot identify particular individuals; it can, however, exclude individuals or place individu-

¹⁰² See 1 Giannelli & Imwinkelried, *supra* note 78, § 17-8; George F. Sensabaugh, Jan Bashinski & Edward T. Blake, *The Laboratory's Role in Investigating Rape*, *Diagnostic Med.*, Mar. 1985, at 4.

¹⁰³ See 1 Giannelli & Imwinkelried, *supra* note 78, § 17-8.

¹⁰⁴ See, e.g., Thomas C. Meyers, *The Role of the Forensic Serologist*, in *Forensic Science and Law* 409, 409 (Cyril H. Wecht & John T. Rago eds., 2006).

als within a percentage of the population that possesses a given type and cannot be excluded as a source of the fluid.¹⁰⁵

The ABO frequencies were derived from well-established, scientifically valid databases. Based on the analysis of more than 70,000 samples, it was observed that approximately 40% of the Caucasian population are Type A, 11% are Type B, 45% are Type O, and 4% are Type AB.¹⁰⁶ For Blacks, 23% are Type A, 22% are Type B, 51% are Type O, and 4% are Type AB.¹⁰⁷ The most crucial element of any conclusion concerning serology is the relevant population that is included by a finding of blood markers in a crime scene sample. Any testimony that misstates the relevant included population violates the scientific basis for conventional serology.

Serologists in these cases often violated that scientific criterion by misstating the included population in a manner that made their findings appear more probative. Most often they claimed the perpetrator was part of some subset of the population that included the defendant, when in fact no part of the population could be excluded. In other cases, they inaccurately narrowed the subset of the population including the defendant. In still additional cases, the serology excluded the defendant, but analysts argued that the results were non-probative or could somehow nevertheless include the defendant. In each of these examples of invalid testimony, the analyst misstated the statistics regarding the included population to make them seem smaller and therefore more inculpatory than they in fact were.

1. Ignoring the Problem of Masking and Quantification

Most of the DNA exonerations involved sexual assault convictions. During the criminal investigations in most of those cases, a rape kit was prepared, which would include swabs taken by doctors

¹⁰⁵ A few courts bar serology results including the defendant, fearing that jurors might misunderstand statistical evidence regarding the population included or deeming such results legally inconclusive. See 1 Giannelli & Imwinkelreid, *supra* note 78, § 17-9.

¹⁰⁶ 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).

¹⁰⁷ *Id.*

from the victim's body.¹⁰⁸ In addition, law enforcement might preserve other crime scene evidence, such as clothing, on which the assailant may have deposited fluids. Sexual assault cases typically involve mixed stains, in which the victim's own genetic markers may often be present and obscure the genetic markers from the assailant.

While modern DNA techniques allow analysts to isolate and amplify miniscule amounts of semen contained in a mixed stain, conventional serology was not capable of doing so. The proportion of semen in the sample could be so small that any material from the semen would not be detected; this is known as the problem of masking and quantification. The victim's own genetic markers could overwhelm—or “mask”—any genetic markers from the semen, making it impossible to detect the blood antigen type of the assailant absent the ability to quantify the semen content of the sample.¹⁰⁹ As Blake put it, because “[s]emen evidence is normally contaminated with vaginal material from the victim,” the interpretation of such evidence “must take into consideration the possible contribution of the victim to the genetic marker pool.”¹¹⁰ This problem was well known in the 1980s, when most of the people later exonerated by DNA testing were convicted.¹¹¹ Quantification techniques later made it possible to reach additional conclusions re-

¹⁰⁸ Fed. Bureau of Investigation, *Handbook of Forensic Sci.* 35–36 (1984) [hereinafter 1984 FBI Handbook].

¹⁰⁹ In other words: “Conventional serology is further limited, in that analysis of mixed-fluid stains in which two or more contributors are involved can mask an individual donor.” National Research Council, *DNA Technology in Forensic Science* 158 (1992).

¹¹⁰ Affidavit of Edward T. Blake, *supra* note 3, at 15.

¹¹¹ The 1984 FBI Handbook explained that in a mixed stain, with both seminal and either urine or vaginal secretions, testing “is more difficult,” and further the donor might be a “‘weak’ secretor” or the “amount of blood group factor present in the semen or saliva” may be “insufficient for reliable grouping tests.” 1984 FBI Handbook, *supra* note 108, at 34; see also *Forensic Sci. Research and Training Ctr., Proceedings of a Forensic Science Symposium on the Analysis of Sexual Assault Evidence* 61 (1983) (describing the situation in which “the group of the victim masks or hides that of the assailant’s semen”); Brian Wraxall & Thomas Fedor, *Oklahoma City Police Department Crime Laboratory Serology Audit, Serological Research Institute, Report 4* (2001) (finding that where the analyst observed ABO typing results consistent with the victim, “the only conclusion that she should correctly draw is that the semen donor could be any male in the population. These interpretation rules were well known and established in 1986.”).

garding mixed-fluid stains in which no antigens foreign to the victim were found. Such techniques were developed in the mid-1980s, but were not widely adopted by crime laboratories until later. None of the cases in the study set with invalid testimony involved the use of techniques for the quantification of semen.

In a case involving a mixed stain in which no ABO blood group substance or enzymes foreign to the victim are detected, any forensic serologist knows that, absent additional information based on quantification of the semen content of the sample, “no potential semen donor can be excluded.”¹¹² Under such circumstances, the failure to inform the jury that 100% of the male population could be included and that none can be excluded is highly misleading. In David Sutherlin’s case, Ronald Enzenauer, of the State of Minnesota, Bureau of Criminal Apprehension, properly explained this phenomenon:

Q. So that [ABO typing] test—you can’t tell anything about the donor because she masks all of those blood groupings?

A. That is correct.

Q. Would there be any blood grouping that that wouldn’t mask?

A. No.¹¹³

The problem of masking and quantification may assist the State to explain why seemingly unfavorable serology evidence is “neutral evidence at worst.”¹¹⁴ Masking and quantification can also be important to the defense, to show why seemingly inculpatory serology evidence is in fact non-probative. In 46 of the invalid serology testimony cases, the analysts described serological results from a mixed stain, yet failed to explain that nothing informative could be said about the potential semen donor population; the serological evidence included 100% of the population. Instead, the analysts testified that the perpetrator was included within some smaller

¹¹² Michael R. Bromwich, Fourth Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room 19 (Jan. 4, 2006), available at <http://www.hpdlabinvestigation.org/reports/060104report.pdf>.

¹¹³ Trial Transcript at 252, 263, *State v. Brian Sutherlin*, No. 4181 (Minn. Dist. Ct. June 7, 1985).

¹¹⁴ *Driscoll v. Delo*, 71 F.3d 701, 708 (8th Cir. 1995).

2009] *Invalid Forensic Science Testimony* 39

percentage of the population that could have produced the semen recovered from the rape victim.

State of California v. Herman Atkins

One example is the Herman Atkins case. There, the victim was a Type A secretor, PGM type 2+1+, as was Atkins. Similarly, the vaginal swab test results disclosed Type A secretor, PGM type 2+1+ material, which was consistent with both the victim and Atkins. James Hall, of the Riverside Crime Laboratory, California Department of Justice, testified as follows:

Q. Based on the information that you received, what kind of conclusions could you tell me about the swab and the blood type of the two persons?

A. Well, the blood type of the swab was the same blood type of the two individuals. That tells me that possibly the semen . . . could be of that blood type combination, or the activity that I detected could have come from the victim herself, or it could have come from a combination of individuals with those blood types.

Q. Do your results exclude the person that you identified as Herman Atkins' blood?

A. No, they don't.

Q. Now, are there certain percentages of the population that have this ABO typing and the PM—what you discovered?

A. PGM.

Q. PGM that you discussed?

A. Yes, there are.¹¹⁵

Hall then stepped down from the stand and made a chart in front of the jury. He wrote the figure 4.4% next to the word "black," and this testimony followed:

Q. It would be 4.4% of the black population?

A. That's right.¹¹⁶

¹¹⁵ Trial Transcript at 233–34, *People v. Herman Atkins*, No. 28832 (Cal. Ct. App. Aug. 22–25, 1988).

¹¹⁶ *Id.* at 234–36.

In the Atkins case, the serologist found nothing foreign to the victim and he made no assessment of the quantity of semen in the mixed body fluid sample. Rather than testifying that no conclusion could be drawn about the genetic profile of the semen donor, Hall presented a statistic: 4.4% of the black population are Type A secretors, PGM type 2+1+, thus excluding more than 95% of the population and including Mr. Atkins as a potential semen donor. One might argue in defense of Mr. Hall that all he did was accurately answer the prosecutor's irrelevant question. But since the analyst knew that the *only* frequency relevant in a sexual assault case is the combined frequency of the potential semen donors, by providing an irrelevant frequency in response to the prosecutor, the testimony misled the jury. National symposia on serology left no scientific doubt about these principles.¹¹⁷

Again, the only probative frequency statistic that is considered by a forensic scientist investigating a sexual assault is the combined frequency of all possible genotypes potentially possessed by the semen source. In those cases where all of the traits detected in the sample could originate from the female body fluids and there is no assessment of the semen dilution, the potential types for the semen source encompass the entire population and no one can be eliminated as a potential semen source.

State of Texas v. Kevin James Byrd

Another example is the case of Kevin James Byrd, in which James Bolding of the City of Houston Police Department Crime Laboratory found no antigens in an examination of a vaginal swab from the rape kit. The victim was a Type B non-secretor, and Byrd was a Type A non-secretor.¹¹⁸ Bolding testified as follows:

A. My conclusion would be that the individuals present or the samples present contained a non-secretor status, that is, we could not determine whether or not they had type A, B, or O blood factor.

¹¹⁷ See supra note 111.

¹¹⁸ Trial Transcript at 164-69, State v. Kevin Byrd, No. 424823 (Minn. Dist. Ct. Aug. 7, 1985).

2009] *Invalid Forensic Science Testimony* 41

Q. Does that mean the victim in the case would have been a non-secretor?

A. That would be the conclusion we would draw.

Q. What would the conclusion you would draw be regarding the suspect or the attacker in this case?

A. That would also mean that the semen donor would also be a non-secretor.¹¹⁹

Bolding testified that 15–20% of the population are non-secretors, and that the defendant was a non-secretor.¹²⁰ However, no quantification was conducted to assess the degree to which semen was present in the sample. As a result, the failure to observe any ABO blood group substances could be due to an inadequate concentration of semen in the extract employed to conduct the ABO typing. The proper statistic, therefore, was that 100% of the population could have been the source of the semen because there was no information to prove that the quantity of semen was adequate to detect blood group substances, even if the semen contained them.

Because the Type O secretor, in which the individual secretes only the H antigen, is the most common ABO type, many of the masking cases involved the common situation in which the victim and the defendant were both Type O secretors and the materials from the rape kit exhibited just the H antigen. In all such cases, absent quantification, 100% of the population could have been the semen donor. Yet in a series of cases, the examiner testified that the defendant, a Type O secretor, was included in some subset of the population that could have committed the crime.

*Commonwealth of Massachusetts v. Neil Miller and
Commonwealth of Massachusetts v. Marvin Mitchell*

In the Neil Miller case, David Brody of the Boston Police Department testified that “[t]he H blood group substance that I found had to be deposited by a Group O individual, a Group O secretor”; additionally, he stated that “[a]pproximately forty-five percent of

¹¹⁹ Id. at 164–65.

¹²⁰ Id. at 165–66.

the population are Group O individuals.”¹²¹ Brody’s testimony is particularly disturbing because in the Marvin Mitchell trial—also included here—he understood well the problem of masking and quantification. At Mitchell’s trial, Brody testified regarding the phenomenon where it assisted the prosecution in attempting to explain why Mitchell, a Type A secretor, was not excluded by a test of the vaginal swab containing H antigens only, which was consistent with the victim, a Type O secretor. There Brody testified:

A. Mr. Mitchell could not be excluded. *No secretor could be excluded from depositing that stain because the stain may have been too diluted or graded [sic] to pick up Mr. Mitchell’s blood type. So I cannot exclude him, but I cannot say that I found the A blood group type. In other words, again no secretor can be totally excluded from the stain.*¹²²

Even in the Mitchell case, Brody did not fully explain the phenomenon of masking; he erroneously implied that as a secretor, Marvin Mitchell was part of some subset of the population that could have been the donor. He never informed the jury that no individual, whether a secretor or non-secretor, could be excluded.

2. Invalid Analysis of Presumptively Exculpatory Serology

In a series of cases, traditional serology evidence strongly supported the defendant’s innocence, but forensic analysts engaged in a series of unsupported and misleading speculations on the stand attempting to discount the evidence of innocence. Sometimes the testimony converted exculpatory serology findings into false inculpatory evidence. These cases typically involved analysts telling the jury that antigens can “selectively degrade” due to bacterial contamination, thus disregarding a result that excluded the defendant. The Supreme Court recently decided a case involving invalid testimony of this sort by an FBI analyst.¹²³

¹²¹ Transcript, N. Miller, *supra* note 64, at 198.

¹²² Trial Transcript at 51–52, *Commonwealth v. Marvin Mitchell*, No. 072574 (Mass. Dist. Ct. Jan. 18, 1990) (emphasis added).

¹²³ See *House v. Bell*, 126 S. Ct. 2064, 2071–72 (2006); Brief for the Innocence Project, Inc. as Amicus Curiae in Supporting Petitioner at 25, *House v. Bell*, 126 S. Ct. 2064 (2006) (No. 04-8990) (noting that “if antigens could selectively ‘vanish,’” and

State of Montana v. Paul Demetri Kordonowy

In the case of Paul Kordonowy, Julie Long of the Montana State Forensic Laboratory performed the testing on the victim's underpants that revealed A antigens, which neither the victim nor Kordonowy possessed; both were Type O secretors.¹²⁴ Rather than testify that the testing excluded Kordonowy, Long testified as follows: "[I]n this case there was a large amount of bacteria, which I noted, and it has been reported that a large amount of bacteria can give you an A substance reading in your analysis because your ABO substances are sugars, and bacteria also produce sugars."¹²⁵ As Edward Blake concluded in his report examining the case, this "bacteria" testimony lacks any scientific foundation:

[T]here is no evidence whatsoever that bacteria produce water soluble ABO antigens of any sort much less ABO antigens of type "A." If this assertion were true, the ABO typing of sexual assault evidence would be inherently unreliable because no scientist could ever know whether or not the ABO antigens detected in vaginal or oral samples were from ubiquitous bacteria or the human being from whom the sample was collected or some other human being contributing a body fluid to the sample. Like the claim that bacteria preferentially destroy spermatozoa, the claim that bacteria preferentially secrete ABO "A" antigens is without scientific basis; and, if true, would undermine the entire scientific foundation for the ABO typing of body fluid evidence.¹²⁶

Indeed, Long not only stated that based on her analysis, Kordonowy could not be excluded,¹²⁷ but went further by affirmatively including Kordonowy. Long stated that Kordonowy fell within the 29% of the population who are Type O secretors and could be the

blood types could mutate from one to another, "then serology would *never* have been a reliable method"); Garrett, *supra* note 55, at 1638.

¹²⁴ Trial Transcript at 355, 369–70, *State v. Paul Demetri Kordonowy*, No. DC-89-013 (Mont. Dist. Ct. Jan. 16–18, 1990) [hereinafter Transcript, Kordonowy].

¹²⁵ *Id.* at 371.

¹²⁶ See Lori E. Hutchinson, Report on *State v. Paul Kordonowy* 3–4 (2003).

¹²⁷ Transcript, Kordonowy, *supra* note 124, at 385.

donor.¹²⁸ That testimony was also invalid. Putting aside her failure to give correct attribution to the A antigen, Long failed to recognize that the H blood group substance was not foreign to the victim and hence cannot be used to limit the population of semen donors.

State of Illinois v. Gary E. Dotson

The case of Gary Dotson, the first person exonerated by post-conviction DNA testing, also included the same type of invalid testimony—in addition to the invalid testimony already described that ignored the problem of masking and quantification. Timothy Dixon of the Illinois Department of Law Enforcement told the jury not to reach any conclusions based on the Type A antigens found in stains in several places on the victim's clothes, antigens that could not have come from the victim or Dotson, who were both Type B. Dixon testified:

The A stain—I can't say the A stain, I can't say that blood is A, I can't say that blood is B, all I can say is that material was blood, and a mixture of—it could be perspiration, could be other body fluids in combination of B and H activity.¹²⁹

He added: “Unfortunately for us there are lots of materials; dust, wood, leather, certain kinds of clothes, different cloth materials, detergents in materials” that could somehow explain the presence of the Type A antigens.¹³⁰

As Blake explained in his report, control testing could be used to test the area outside the stain to assess whether the result was due to contamination. If it were actually the case that contamination could never be ruled out, then “ABO typing of biological samples” would have always been an “inherently unreliable” type of analysis.¹³¹

¹²⁸ *Id.* at 386. Long agreed with the statement that “we cannot rule out Mr. Kordonowy because of the presence of the H Substance,” and then added that 29% of the population are Type O secretors. *Id.* at 383, 385–86.

¹²⁹ Trial Transcript at 359, *State of Illinois v. Gary E. Dotson*, No. P.C. 4333 (Ill. Cir. Ct. July 25, 1985).

¹³⁰ *Id.* at 368.

¹³¹ Affidavit of Edward T. Blake, *supra* note 3, at 23.

2009]

Invalid Forensic Science Testimony

45

State of West Virginia v. Gerald Davis
State of West Virginia v. Dewey Davis

Similarly, during the trials of Gerald and Dewey Davis, Fred Zain of the West Virginia Department of Public Safety, Criminal Investigation Bureau, gave misleading testimony. In Gerald Davis's trial, Zain explained how bacteria could account for the presence of Type A antigens where the victim and Gerald Davis were both Type O secretors, stating, "Bacterial contamination can give you what is called false positives and give you blood types separate and aside from what you're truly identifying."¹³² Where the testing excluded Gerald Davis, Zain instead claimed that by ignoring the Type A finding, one should conclude that Davis was included in the group of "around the realm of 7 percent of the general population of West Virginia" who could have committed the crime.¹³³

The American Society of Crime Laboratory Directors Laboratory Accreditation Board ("ASCLD/LAB") Investigation Report of Zain's work conducted in 1993 concluded that finding "an ABO type foreign to both the victim and defendant . . . would normally be interpreted as excluding defendant as the semen donor," and that there was "no satisfactory foundation" for the conclusion in the Gerald and Dewey Davis trials that resulted in the "ABO mismatch" being "dismissed as bacterial contamination."¹³⁴

¹³² Trial Transcript at 259, State v. Gerald Davis, No. 86-F-152 (W. Va. Cir. Ct. July 21, 1986) [hereinafter Transcript, G. Davis]. Or as Zain put it in Dewey Davis's trial: "[A]nytime a body fluid leaves an individual's body, there is an automatic contamination of whatever the body fluid might be up to and including the time that that stain becomes dried." Trial Transcript at 238, State v. Dewey Davis, No. 86-F-153 (W. Va. Cir. Ct. Mar. 9, 1987). Zain added:

You have foods that once you open a can—I'm sure most of the ladies on the jury have done some canning at one time or another. Once you open that up, you have to either keep it in a refrigerator or you have to have it in sort of a preservative to keep bacterial activity from happening. . . . Blood and body fluids are the same thing.

Id. at 239.

¹³³ Transcript, G. Davis, supra note 132, at 249.

¹³⁴ See James J. McNamara, Ronald R. Linhart, ASCLD/LAB Investigation Report, West Virginia State Police Crime Laboratory, Serology Division 8 (1993); see also In re W. Va. Crime Lab, Serology Div, 190 W.Va. 321 (1993).

State of Indiana v. Jerry Watkins

Several other similar examples are included in the data set. In the case of Jerry Watkins, the victim was a Type A secretor, Watkins was a Type O secretor, and yet the swabs displayed the A, B, and H blood group substances.¹³⁵ Forensic analyst Carol Kohlman of the Indianapolis and Marion County Forensic Services Agency was asked, “Is there anything from your results that would allow you to exclude the possibility of Jerry Watkins being the semen donor?” Despite the presence of the B blood group substance, which was foreign to both the victim and Watkins, she answered, “No sir.” She gave another similar explanation: “I was suspecting bacterial contamination as part of the problem or as a possible explanation”¹³⁶ The defense attorney did question Kohlman regarding this surprising opinion on cross-examination, asking whether bacteria of such a kind were observed, asking, “Did you do any cultures?” She responded, “No sir, we do not do cultures in our laboratory.”¹³⁷ No effort was made to support the bald conjecture, nor was the valid result—that the defendant should have been excluded—ever presented.

3. Additional Invalid Use of Population Statistics

In addition to the large set of cases involving invalid testimony that ignores the problem of masking and quantification, several other examiners misstated or miscalculated population statistics. In a series of cases, serologists divided the relevant statistic in half, claiming that men constitute half of the population and only a male could have deposited semen at the scene. It is logically incorrect to divide a frequency in half when it is understood at the beginning that the relevant population is males, because only males produce semen. The population statistics regarding these blood group substances are identical for both sexes. For example, approximately 40% of both Caucasian men and women are ABO Type A. It is incorrect to divide that figure in half and claim that only 20% of men are Type A.

¹³⁵ Trial Transcript at 961, 977, 988–89, *State v. Jerry Watkins*, No. 30D01-8603-CF-005 (Ind. Super. Ct.) (date unknown).

¹³⁶ *Id.* at 999.

¹³⁷ *Id.* at 1025.

In the Mark Bravo case, Richard W. Catalani of the Los Angeles County Sheriff's Crime Lab stated that although 3% of the population was PGM type 2-1+, that figure should be divided in half to eliminate females. That testimony provided the jury with the incorrect figure that 1.5% of the male population could have been the semen donor.¹³⁸ In the Perry Mitchell case, John Barron of the South Carolina Law Enforcement Division, noting that H antigens were observed and that 35% of the population are Type O secretors, testified, "You would probably have to also cut that by another 50% because we're dealing with males." When the defense counsel questioned why population statistics are not the same for men and women, Barron responded, "[T]here is a difference in regard to semen."¹³⁹ Citing the same false rationale, Fred Zain similarly divided statistics in half in the Dewey Davis and Glen Woodall cases, as did Julie Long in the Chester Bauer case.

To summarize, in each of these examples of invalid forensic science testimony, the analyst failed to accurately provide the relevant statistic regarding the included population. These analysts instead offered invalid, reduced frequencies (a rarer event) that appeared to further inculpate the defendant.

B. Invalid Microscopic Hair Comparison Testimony

Sixty-five of the trials examined involved microscopic hair comparison analysis. Of those, 25—or 38%—had invalid hair comparison testimony. Most (18) of these cases involved invalid individualizing claims.

Forensic hair microscopy involves the side-by-side comparison under a microscope of head and pubic hairs found at a crime scene with dozens of head and pubic hairs plucked and combed from the scalp and pubis of the victim and suspect(s).¹⁴⁰ Hair examination has long been an important part of police investigations, because

¹³⁸ Trial Transcript at 267-68, *People v. Mark Diaz Bravo*, No. Va. 003313 (Cal. Super. Ct. Sept. 27, 1990).

¹³⁹ Trial Transcript at 225, *State v. Perry Mitchell*, 83-GS-32-479 (S.C. Cir. Ct. Jan. 19, 1984).

¹⁴⁰ Richard E. Bisbing, *The Forensic Identification and Association of Human Hair*, in 1 *Forensic Science Handbook* 390, 414-21 (Richard Saferstein ed., 2002).

hairs are so commonly and readily transferred to skin or clothing.¹⁴¹ A head hair found on the sheets of a rape victim is a “questioned” hair; the numerous hairs plucked and combed from the victim and suspect are “reference exemplars.”

Forensic hair evidence has increasingly been scrutinized due to studies indicating high error rates.¹⁴² Examiners commonly distinguish human from animal hair, opine on the racial group from which the questioned hairs originated, determine from which part of the body the hair originated, and then testify that the hairs have microscopic characteristics that are similar or dissimilar to the exemplar samples.¹⁴³ This study is not concerned with the analyst’s ability to distinguish hair of different species, races, or parts of the body. This study is concerned with testimony which attempts to declare a positive association between a questioned hair from a crime scene with a set of exemplars from a suspect.

Not only was forensic human hair comparison frequently relied upon in criminal cases at the time relevant to this study because hairs are easily left at a crime scene, but also because there is considerable variation in the microscopic characteristics of hairs coming from different people. There exists significant intrapersonal variation among the hairs from a single individual’s head, and as a result, the competent examiner will collect perhaps 100 hairs taken from different regions of the scalp and then select a representative subset of perhaps 20 hairs to be compared with the questioned head hair. There are many different general or “class” characteristics of hair that can be microscopically examined and compared along the length of the hair. Some of the general characteristics are color, structure, and cuticular traits. Although no single class characteristic is very unusual, 20 representative hairs possessing a range

¹⁴¹ Edward J. Imwinkelried, *Forensic Hair Analysis: The Case Against the Underemployment of Scientific Evidence*, 39 Wash. & Lee L. Rev. 41, 42–43 (1982).

¹⁴² *Id.* at 44–45 (describing a Law Enforcement Assistance Administration study); D. Michael Risinger & Michael J. Saks, *Rationality, Research and Leviathan: Law Enforcement-Sponsored Research and the Criminal Process*, 2003 Mich. St. L. Rev. 1023, 1049–50 (deriving an 11.5% error rate for visual hair comparison from FBI study data); see also Clive Stafford Smith & Patrick D. Goodman, *Forensic Hair Comparison Analysis: Nineteenth Century Science or Twentieth Century Snake Oil?*, 27 Colum. Hum. Rts. L. Rev. 227, 242–45 (1996).

¹⁴³ Bisbing, *supra* note 140, at 418–19.

of characteristics that are similar to the questioned hair would be forensically significant.¹⁴⁴

As the FBI noted in its 1984 handbook, microscopic hair examination is “[n]ot positive evidence.”¹⁴⁵ The National Research Council has explained:

Although hair examiners can associate a hair with racial characteristics and body source (trunk, head, or pubic area) the variations among hairs on a given person make definitive association of a single hair with an individual problematic. The microscopic comparison of hairs is also subjective and can lead to differences of opinion among equally qualified experts.¹⁴⁶

Where even qualified examiners may disagree about whether hairs could come from the same source, hair evidence cannot be individualized based on microscopic analysis.¹⁴⁷ Consequently, the field adopted standards that the strongest statement of association that can be made is a statement that hairs in question are “consistent” with the defendant’s or “could have” come from the defendant. The 1985 International Symposium on Forensic Hair Comparisons (“1985 Symposium”) adopted these standards.¹⁴⁸

Thus, hair examiners following those standards may testify that two samples are visually or microscopically similar. Though “valid” in the sense used here—because such testimony does not misstate empirical data or incorrectly claim empirical support—such a conclusion is not highly probative. As Richard Bisbing has put it, such

¹⁴⁴ Moreover, in some cases, there may be an artificial treatment which can be of value. For example, if a suspect had brown hair, dyed it green, and then let it grow out, and the length that was brown versus the length that was green would be approximately the same between the questioned hair and the exemplars; the similarity would be significant. *Id.* at 410.

¹⁴⁵ 1984 FBI Handbook, *supra* note 108, at 37; see Bisbing, *supra* note 140, at 419 (“[O]ne can never say with absolute certainty that a particular hair originated from one individual to the exclusion of all others.”).

¹⁴⁶ Comm. on DNA Tech. in Forensic Sci. Nat’l Research Council, DNA Technology in Forensic Science 158 (1992).

¹⁴⁷ This is an important difference between serology and hair microscopy. In the former, it is expected that all competent analysts will agree on whether two samples have the same blood group substances.

¹⁴⁸ See FBI, Lab. Div., Proceedings of the International Symposium on Forensic Hair Comparisons 108 (1985) [hereinafter 1985 Symposium]; discussion *infra* Subsection II.B.1.

evidence is “by necessity, equivocal, that is ‘could be evidence.’”¹⁴⁹ Evidentiary questions that are not addressed here are raised by such testimony. Courts should ask whether a finding that hairs are “similar” or “consistent” has sufficient probative value to be admissible, or causes undue prejudice due to the ambiguity of the terms “similar” and “consistent.” Important questions remain whether such subjective analysis and such imprecise conclusions would today satisfy *Daubert* requirements for admissibility of expert testimony.¹⁵⁰ The NAS report was emphatic that “[n]o scientifically accepted statistics exist about the frequency with which particular characteristics of hair are distributed in the population.”¹⁵¹ The report added that, “[t]here appear to be no uniform standards on the numbers of features on which hairs must agree before an examiner may declare a ‘match.’”¹⁵² Linking these defects with the problem of invalid terminology used in reports and testimony, the NAS report explained that “[t]his illustrates not only the imprecision of microscopic hair analyses, but also the problem with using imprecise reporting terminology such as ‘associated with,’ which is not clearly defined and which can be misunderstood to imply individualization.”¹⁵³ A range of other forensic disciplines lacking empirical data—such as bite mark comparison and handwriting comparison—raise these questions and also may not survive proper *Daubert* scrutiny.¹⁵⁴ However, for the purposes of this study, which does not reach such questions, testimony solely observing a similarity, while imprecise and potentially misleading, is deemed valid, because it does not render a conclusion contrary to

¹⁴⁹ Richard E. Bisbing, Forensic Hair Comparisons: Guidelines, Standards, Protocols, Quality Assurance and Enforcement, Presentation to NAS, April 24, 2007, <http://www7.nationalacademies.org/stl/April%20Forensic%20Bisbing.pdf>.

¹⁵⁰ See, e.g., 2 Giannelli & Imwinkelried, *supra* note 78, § 24-3; Imwinkelried, *supra* note 141; Smith & Goodman, *supra* note 142, at 283–90. But see Houck et al., The Science of Forensic Hair Comparisons and the Admissibility of Hair Comparison Evidence: Frye and Daubert Considered, Mar. 2004, <http://www.modernmicroscopy.com/main.asp?article=36,12Mar2004>.

¹⁵¹ See Strengthening Forensic Science, *supra* note 12, at 5-25.

¹⁵² *Id.*

¹⁵³ *Id.*

¹⁵⁴ See, e.g., Saks & Koehler, *supra* note 50, at 218.

underlying empirical data or claiming support based on nonexistent data.¹⁵⁵

DNA testing of the mitochondria, or when the hair roots are present, of the nucleus, has now supplanted microscopic hair comparison in many cases. In six exonerees' cases, for example, the analyst identified hairs as consistent with the defendant at trial, but mitochondrial or other DNA analysis later determined that those same hairs originated from a person other than the convicted defendant.¹⁵⁶

1. Invalid Probability Testimony

Where an examiner can only opine whether hairs are similar or consistent, forensic hair comparison experts also advise that an examiner should not make "any statements about the probability that a specific hair could have come from someone other than the person to which it was associated."¹⁵⁷ The 1985 Symposium noted the possibility of a "coincidental match," and therefore cautioned that there was a need for "further research" on frequency before probability statements can be used when describing a hair compari-

¹⁵⁵ See *id.* at 216 (recommending that until empirical research is done to support certain forensic disciplines, for the present, "criminalists should report [a] finding with the appropriate clarity and restraint. For example, they could explain that a conclusion that two patterns are consistent (or a match) does not require a conclusion that the patterns share a common source" and noting that "there is no scientific justification" for speculation regarding the likelihood that the patterns share a common source).

¹⁵⁶ The cases are those of Richard Alexander, William Dedge, George Rodriguez, Drew Whitley, Stephen Avery, and Anthony Hicks. See Innocence Project, <http://www.innocenceproject.org/know/Browse-Profiles.php>.

¹⁵⁷ See Max M. Houck et al., *The Science of Forensic Hair Comparisons and the Admissibility of Hair Comparison Evidence: Frye and Daubert Considered*, *Mod. Microscopy J.* 5 (Mar. 2, 2004), available at <http://www.modernmicroscopy.com/main.asp?article=36&page=5&searchkeys=Houck> ("All of these numbers notwithstanding, to attempt to derive a population frequency of traits or to determine how likely it may be to encounter a given hair in a given population is fraught with complexity. Most experts . . . do not feel comfortable with any statements about the probability that a specific hair could have come from someone other than the person to which it was associated. The authors agree with that approach. The justification for that reluctance is based on the complexity of the probability question, difficulty choosing a population to which to assign the probability, the lack of sufficient data where that question was addressed, and court decisions excluding such statements of probability in the past." (citations omitted)).

son.¹⁵⁸ No such systematic efforts to research the frequency with which particular microscopic features occur in any population have been conducted. Thus, there is not and never has been any statistical basis for hair comparison.¹⁵⁹

After all, examiners lack empirical data on the frequency of any of the general or “class” characteristics, and lack data as to the extent to which there is a correlation between different class characteristics. Without frequencies (in contrast to the ABO blood group system), hair examiners are also precluded from expressing to the jury a probability that hairs recovered at the crime scene came from the defendant. A probability can ordinarily be determined in one of two ways: theoretically or empirically. A theoretical probability requires a well understood situation under demonstrable assumptions. If one rolls a six-sided die, assuming that each of the six sides is equally likely to show up on top, the theoretical probability of any particular side showing up in a single roll of the die is 1/6. Given what is known about hair, hair analysis itself does not lend itself to theoretical probabilities. Empirical probabilities, on the other hand, are gained from a large set of data expressed as: total number of times the outcome occurred divided by total number of instances examined. Since there are no adequate sets of data for the occurrence of general hair characteristics, the examiner cannot present an empirical probability.

¹⁵⁸ 1985 Symposium, *supra* note 148, at 110; see also James Robertson & Colin Graeme Girdwood Aitken, *The Value of Microscopic Features in the Examination of Human Head Hairs: Analysis of Comments Contained in Questionnaire Returns*, 31 *J. Forensic Sci.* 563, 568 (1986) (“There is a clear need for an extensive research program to evaluate the microscopic features of hair from a forensic science standpoint . . .”). The “general opinion” among experts in the mid-1980s, for example, was that “Gaudette’s study is only an initial step toward determining the significance of hair comparison evidence,” and that “[f]orensic experts, including those employed by the FBI, still recommend that hair examiners use cautious statements when asked to give an opinion on whether the origin of an unknown hair and of a representative sample is the same.” Dalva Moellenberg, *Splitting Hairs in Criminal Trials: Admissibility of Hair Comparison Probability Estimates*, 1984 *Ariz. St. L.J.* 521, 536.

¹⁵⁹ See Richard E. Bisbing et al., *Peer Review Report: Montana v. Jimmy Ray Bromgard 2* [hereinafter *Peer Review Report*] (“[T]here is not—and never was—a well established probability theory for hair comparison.”).

2009]

Invalid Forensic Science Testimony

53

State of Montana v. Jimmy Ray Bromgard

Nevertheless, several analysts in these exonerees' cases violated that criterion and bolstered their testimony by making invalid statements regarding supposed probabilities. Arnold Melnikoff, Director of the Montana State Crime Laboratory, testified regarding probabilities in a series of cases. Though there is not, and never has been, any statistical basis for hair comparison, he would simply fabricate frequencies and probabilities based on his own undocumented estimates, rather than any reliable empirical data. He would then go even farther and multiply his made up probabilities by the number of hairs "matched" from different parts of the body, as if each represented independent events. In the case of Jimmy Ray Bromgard, Melnikoff testified that "[t]he hair from the blanket on the left matches all the characteristics of the known pubic hair from James Bromgard on the right, and they almost look like one hair."¹⁶⁰ He then explained the probability of such a "match":

Well there are actually two mutually exclusive events because they come from different areas of the body, and their characteristics are not necessarily the same. So if you find both head and pubic hair there you have one chance in a hundred for the head hair matching a particular individual and one chance in a hundred for the pubic hair. If you find both it's a multiplying effect, it would be one chance in 10,000, it's the same as two dice, if you throw one dice with a one, one chance out of six; if you throw another dice with a one, it's one chance out of six, you multiply the odds together. You do the same in this case so, it's one times one hundred, times one, times one hundred, and you get one in 10,000.¹⁶¹

State of Montana v. Chester Bauer

Arnold Melnikoff testified similarly in Chester Bauer's case, stating:

¹⁶⁰ Trial Transcript at 231, *State v. Jimmy Ray Bromgard*, No. 88108 (Minn. Dist. Ct. Nov. 16, 1987) [hereinafter Transcript, Bromgard].

¹⁶¹ *Id.* at 237-38.

To have them both match, it would be the multiplication of both factors so as an approximately [sic] using that 1 out of 100, you come out with a number like 1 chance in 10,000. Multiply 100 x 100. It becomes a very highly improbable situation that both events would occur, that you could not distinguish the head hair and the pubic hair from two individuals.¹⁶²

And he testified the same way in Paul Kordonowy's case, again claiming that hairs from different parts of the body are "independent events" to which he attached numerical probabilities which he then multiplied.¹⁶³ Not only did Mr. Melnikoff not have any data to support the number 1/100, he also had no information to conclude that the color and class characteristics of head hair and pubic hair are independent so that their frequencies can be multiplied.

The probabilities of two events can only be multiplied if the events are statistically independent; that is to say that the outcome of one event does not influence the outcome of the other event. Melnikoff multiplied the two probabilities without it first being scientifically proven that the events are statistically independent. There is no published research on the question of statistical independence for head and pubic hair. Indeed, on the contrary, a person with dark-colored scalp hair may be more likely than a person chosen at random to have dark colored pubic hair.

State of Oklahoma v. Timothy Edward Durham

In Timothy Durham's case, Carol English Cox of the Tulsa Police Laboratory opined that the particular reddish-yellow hue of his hair and the questioned hair were only found in "about 5 percent of the population."¹⁶⁴ Cox did not provide scientific support for that statistic, nor could she do so.

¹⁶² Trial Transcript at 250, *State v. Chester Bauer*, No. 83-CR-27 (Mont. Dist. Ct. July 16, 1983).

¹⁶³ *Id.* at 251.

¹⁶⁴ Trial Transcript at 385, *State v. Timothy Edward Durham*, No. CF-91-4922 (Okla. Dist. Ct. Mar. 9, 1993).

2. Exaggeration of Probative Value of Hair Association Without Using Numbers

The 1985 Symposium established standards regarding conclusions that analysts may reach regarding association of questioned and exemplar hairs. Statements of association may consist of conclusions that a hair “could have” come from an individual or “is consistent” with an individual’s hair, or “could not have” come from an individual and “is not consistent” with an individual’s hair; “is consistent” is the strongest statement of association permitted.¹⁶⁵ Many of the experts testifying in the trials studied here described and adhered to those standards with care. Nevertheless, these trials are also replete with examples in which analysts expressed far greater certainty that hairs came from a particular defendant. Testimony used to convey strong association ranged from “more likely than not” that questioned hairs came from the defendant, to naming a “quite large” probability that the questioned hairs came from the defendant, to stating that hairs in fact did come from the defendant.¹⁶⁶ Each of these terms implies a known probability, which, as explained above, does not—and, in the absence of empirical data, cannot—exist. Such testimony or analysis violates the scientific criterion that statements of frequency or probability must be supported by empirical population data.

¹⁶⁵ See 1985 Symposium, *supra* note 144, at 108–09.

¹⁶⁶ Some examiners in the cases studied declared a “match” between the questioned and known hairs. On its face, the word “match” may not seem invalid because it need not imply individualization. For instance, the co-authors “match” in appearance at the most basic level—we each have two arms and two legs. It all depends upon what is meant by “match.” In criminal jury trials, “match” was commonly used to mean individualization. The most frequent use of “match” is when an analyst opines that a crime scene fingerprint “matches” the suspect’s; there the examiner is attempting to communicate uniqueness. For that reason, many hair examiners would shy away from using the word “match” with reference to hair, particularly since there is rarely a one-to-one correspondence between the questioned hair and a single known hair. Rather, in most cases of positive hair association, the characteristics exhibited in the questioned hair fit within the *range* of characteristics reflected in the set of hair exemplars. The word “match” misleadingly implies to jurors that the analyst observed a perfect and complete consistency between only two hairs. Yet the 1985 Symposium did not take a clear position on “match” and some analysts still use the word. Thus, if use of “match” was the only transgression in an expert’s testimony, this study did not conclude the case involved invalid testimony. In the cases deemed invalid, the examiners went further to amplify the probative value of the evidence.

State of Illinois v. Willie L. Rainge, Kenneth E. Adams, and
Dennis Williams

One example of an expression of an invalid degree of non-numerical associative certainty is in the “Ford Heights Four” case. In the trial of Kenneth Adams, William Rainge, and Dennis Williams, Michael Podlecki of the Illinois Department of Law Enforcement Crime Lab testified, while displaying side-by-side photomicrograph comparisons,¹⁶⁷ regarding two hairs found in the trunk of Dennis Williams’s car: “I couldn’t distinguish if I was looking almost at two hairs. They looked just like one.”¹⁶⁸ He added:

What I saw was a hair like this where I could see all the pieces like a puzzle. Like in the previous hair. A line would just fit in. In other words it wasn’t a type of situation of sitting down and looking for it. Just like if you dropped two dollar bills and you see two dollar bills on the floor. You see two one dollar bills. It’s obvious. And that’s how it looked there.¹⁶⁹

Yet when asked to state his ultimate conclusions regarding the exhibits containing the hair evidence, Podlecki opined not that the hairs were identical, but that they were “similar in color and characteristics.”¹⁷⁰ Where the evidence supported only a conclusion that questioned hairs exhibited a range of characteristics found in the exemplar hairs, it was quite misleading to describe the hairs as identical, and then use a further misleading visual display to convey identity to the jury.

¹⁶⁷ Using such visual displays was itself considered improper because they could easily mislead the jury. See 1985 Symposium, *supra* note 144, at 112 (“The Subcommittee is strongly opposed to showing side by side photomicrographs It can be highly inflammatory to a jury.”).

¹⁶⁸ Trial Transcript at 2223, *People v. Willie L. Rainge*, No. 78-I6-5186 (Ill. Cir. Ct. Oct. 10, 1978) [hereinafter Transcript, Rainge].

¹⁶⁹ *Id.* at 2226.

¹⁷⁰ *Id.* at 2227. Podlecki had earlier explained that he had received sets of exemplar hairs, and that in order to conclude that hairs were “similar,” he would have to find a less than 0.1% difference between them. *Id.* at 2208, 2211–12.

2009]

Invalid Forensic Science Testimony

57

State of Oklahoma v. Calvin Scott

Another example is from Calvin Scott's case, in which Claud Berry of the Oklahoma State Bureau of Investigation testified as follows:

Q. Do you know whether or not, Mr. Berry, there have ever been any studies done as to the probabilities of finding another person with hair like ours, or—

A. Well there is one gentleman out of Canada, his name is B.D. Goday [sic], he made a study. He's the only one that has made a study that's been published, and he has found that head hair, one person in forty-five hundred would have a chance of—in other words, identification of one hair to—I mean, one person in forty-five hundred may have features of hair comparison in head hair. Now one in eight hundred in pubic hairs. That's his results. That's the only one I have been able to find who has ever come up with any results with figures. Others have made statements on theory, but they haven't made any practice, or made any study.

Q. Would he have given, or would there be any number type odds to the probability of the hair found on May Ann Fulsom's bottom sheet and the hair, unknown hair found in her pubic combings, both belonging to anyone other than the defendant, Calvin Scott?

A. His hair, I would say this: his studies were made on caucasian hair, I believe. In this case having two hairs identified, two hairs of different kind, I mean, head hair from one person would be quite large, I would say, I would not give a figure. It would be quite large.¹⁷¹

Going beyond expressing a high likelihood of association, Oklahoma City Police Department analyst Joyce Gilchrist expressed a definitive association in the Curtis McCarty case. Gilchrist concluded her direct examination at McCarty's first trial by stating her opinion “[t]hat he was in fact there.”¹⁷² Similarly, in the Larry Pe-

¹⁷¹ Trial Transcript at 47–48, State v. Calvin Lee Scott, No. CRF 83-74 (Okla. Dist. Ct. Sept. 19, 1983).

¹⁷² Trial Transcript at 177, State v. Curtis Edward McCarty, No. CRF-85-02637 (Okla. Dist. Ct. Mar. 24, 1986); see also *infra* Section II.F.

terson case, Gail Tighe of the New Jersey State Police Laboratory agreed that all of the questioned hairs examined were identified as either “belonging” to the victim or to Peterson.¹⁷³

Some examiners did not provide numerical statements regarding probability, but made other efforts to describe the probability of finding comparable hairs, or to describe a particular characteristic as “rare” or “uncommon” without the requisite reliable database from which to draw such inferences.

Commonwealth of Kentucky v. William Gregory

Analysts made conclusions regarding probability based on claims that the hairs had supposedly unusual or unique features. Such claims are unsupported where empirical data regarding the frequency of microscopic features in hair is lacking. During William Gregory’s trial, Dawn Katz of the Kentucky State Police Jefferson Regional Crime Lab testified that the hairs “more than likely” belonged to Gregory.¹⁷⁴ In part, this was based on a finding of “ovoid bodies” in the hairs, which she called “kind of an unusual characteristic.”¹⁷⁵ Katz explained:

A. I told you, there is no statistics [sic] on this. I can tell you this is the first time I have ever had a negroid origin hair that has not had a medulla in it.

Q. What percentage of people have ovoid bodies in them?

A. This is probably the first time I have ever seen an ovoid body in a human hair. I have seen them in cattle hair before.¹⁷⁶

Katz added:

I mean, you have to compare hairs from brothers and sisters that have the same genetic background and carried a lot of the same genetic characteristics from the same parents, you might run into very similar characteristics in two people. But, in general, you

¹⁷³ Trial Transcript at 152, *State v. Larry L. Peterson*, A-3034-89T4 (N.J. Super. Ct. Mar. 6, 1989).

¹⁷⁴ Trial Transcript at 221, 246, *Commonwealth v. William Gregory* (Ky. Cir. Ct. Aug. 1993).

¹⁷⁵ *Id.* at 233.

¹⁷⁶ *Id.* at 251.

2009]

Invalid Forensic Science Testimony

59

wouldn't see that kind of an overlap in two people you would just pick off the street.¹⁷⁷

This testimony was invalid. In addition to the invalid use of probability—claiming that the hair “more than likely” originated from Gregory—Katz testified that a characteristic was “unusual” based on no extant population data. Indeed, she admitted that “there is no statistics [sic] on this.” Katz further embellished that she had never seen such an “unusual” feature before except in “cattle hair.” Finally, Katz implied that only siblings would share such characteristics, again without any data to support such a claim.

The testimony in the Gregory case not only included invalid statements concerning probability, but the analyst made additional claims about particular characteristics based on undocumented experience. Several other examiners buttressed their conclusions by claiming that never in their careers had they seen either special characteristics or more generally, exemplars from any two different people that “matched.” Michael Blair's case involved similar testimony by analyst Charles Linch, then a consultant and formerly of the Southwestern Institute of Forensic Sciences in Dallas.¹⁷⁸ He testified that he had “never seen a Caucasian or Mongoloid hair that was opaque like that,” referring to opaque features he described within the hairs, and added, “I haven't seen a hair like that before. Not a human hair.”¹⁷⁹

In his deposition for the Wilton Dedge case, David Jernigan of the Florida Department of Law Enforcement stated: “Out of all the pubic hairs that I have examined in the laboratory, I have never found two samples, two known samples to match in their microscopic characteristics.”¹⁸⁰

In the trial of Kevin Richardson and Kharey Wise in the so-called “Central Park Jogger” case, retired New York City Police Department Detective Nicholas Petraco was asked whether it was possible that the hairs found did not come from Richardson, but

¹⁷⁷ Id. at 249.

¹⁷⁸ Trial Transcript at 691–92, *State v. Michael Nawe Blair*, No. CRB 21,1152 (Tex. Dist. Ct. Sept. 12, 1994).

¹⁷⁹ Id. at 750–51.

¹⁸⁰ Deposition Transcript at 43, *State v. Wilton Allen Dedge*, No. 82-135-CF-A (Fla. Cir. Ct. June 7, 1984).

rather some unknown person. He responded that although it was possible “[i]n a sense,” it was unlikely:

I’ve looked at thousands of hair standards over the course of my work and I haven’t seen any that have the same range of physical characteristics yet. But I really haven’t looked at them in the sense of exclude one from the other. But I have in fact looked at thousands of standards and haven’t seen two that matched exactly.¹⁸¹

Such testimony exaggerates the probative value of the hair association. As Detective Petraco, who did explain that hairs are not unique, indicated in passing, one would not expect an examiner to have difficulty distinguishing between thousands of standards known to have been obtained from different people. When an analyst claims that in a thousand cases he has never had two reference samples that were not microscopically distinguishable, the rareness of that event is not transferable to estimating the rareness of the association between a questioned hair and a set of exemplars from a known suspect.¹⁸² In a typical sexual assault case the victim is female and the perpetrator is male. Thus, the two sets of reference hairs come from a female and a male. Moreover, the analyst knows in advance that the two sets of samples came from two different people—a rape victim and a suspect. Not only would an analyst be predisposed to differentiate the two sets from one another, but hair length alone can often distinguish such sets of reference samples. Since the analyst’s experience in comparing reference samples to other reference samples answers a very different question than the one that is material to a criminal case, such testimony is misleading.

¹⁸¹ Trial Transcript at 2838, *People v. Kharey Wise*, No. 4762/89 (N.Y. Sup. Ct. Nov. 13, 1990).

¹⁸² See Saks & Koehler, *supra* note 46, at 212–13. Distinguishing one set of exemplars from another when the analyst knows a priori that they originate from two people also introduces context bias. It is a much easier task than comparing a single hair of unknown origin with a collection of hairs taken from a suspect. Moreover, a majority of the hair cases are sexual assaults where in which the victim is a female and the perpetrator is a male. Since one characteristic used to compare hair is length, the question needs to be asked whether in general, hair is more easily distinguishable between men and women. Finally, the hair analyst was limiting his experience to comparing standards in a single case. The analyst was not making inter-case comparisons.

d. Commonwealth of Pennsylvania v. Drew Whitley

Several cases also involved comparisons made where the analyst admitted that the questioned hairs were in fact unsuitable for comparison. For example, in the Drew Whitley case, Dorothy Menges of the Allegheny County Crime Laboratory testified, “Because these hair fragments were so small, I could not make the statement that they were microscopically consistent, but I did see so many overlapping characteristics within the questioned hairs and the standard hairs that I want to make some kind of statement as to their similarities.”¹⁸³

Menges then reversed course and testified, “I found no inconsistencies. Based on what I am basing my comparing on, yes, they are consistent.”¹⁸⁴ After making those statements, she stated: “I wouldn’t go that far to say they were microscopically consistent.”¹⁸⁵ Those equivocations were deeply misleading and unsupported where the fragments were conceded as unsuitable for comparison.

Each of the above examples also suggests a related question: under what circumstances will an examiner conclude that the hair evidence excludes a criminal defendant? Earlier this Part described serology cases in which the analyst failed to exclude. In hair comparison cases, several examiners noted differences but nevertheless concluded that in their estimation, they were not sufficiently “material” so as not to find the hairs to be “similar.”

Some of these cases involved testimony in which experts admitted that the samples possessed manifest differences, but invalidly told the jury that it would be impossible to exclude any person. An example is the case of James Waller, in which Timothy C. Fallon of the Dallas County Institute of Forensic Sciences testified that he “found three dark hair fragments with negroid characteristics that were different from the head and pubic hair of James Douglas Waller.”¹⁸⁶

To conclude that the hair possessed characteristics that could be described as “different,” however, was not sufficient to induce

¹⁸³ Trial Transcript at 898–99, *Commonwealth v. Drew Whitley*, No. CC-8902462 (Pa. Ct. Com. Pl. July 21, 1989) [hereinafter Transcript, Whitley].

¹⁸⁴ *Id.* at 935.

¹⁸⁵ *Id.*

¹⁸⁶ Trial Transcript at 190, *State v. James Douglas Waller*, No. F-82-91294-TK (Tex. Crim. Dist. Ct. Jan. 31, 1983).

Fallon to tell the jury that Waller could not have been the source of that hair. Fallon relied upon speculation, stating that the hair

could have come from another area of the body that we did not take a sample from. It could be that the sample that was the known standard sample that we took from the Defendant was not a large enough random sample to take in all the different characteristics that his hair possessed.¹⁸⁷

Fallon then explained, “if you wanted to say that this hair did not come from this individual, you would have to check it against every hair to be positive that it did not come from that individual.”¹⁸⁸ Fallon told the jury, in effect, that it would be an impossible task to conclude with certainty that a defendant was not the source of a specific hair. He agreed that one would “practically have to denude a person to make a proper comparison.”¹⁸⁹ There was no suggestion that a similar effort should be made for a “proper comparison” permitting an analyst to say that a hair could have come from a defendant.

Similarly, in the case of Habib Abdal (named Vincent Jenkins at the time of trial), Michael R. Krajewski of the Central Police Services Laboratory in Erie County, New York, testified that he “could not make a positive comparison.”¹⁹⁰ By that, he meant that “the hairs—hair samples were distinctively different,” and he explained several key differences.¹⁹¹ He added, “It’s not unusual to have different hairs come from the same person.”¹⁹² On redirect, he explained that even if the exemplar hairs from Abdal did not match the questioned hairs, other hairs of his might match. His hair could have changed over time, or other hairs on his body might

¹⁸⁷ *Id.* at 194.

¹⁸⁸ *Id.* at 195.

¹⁸⁹ *Id.* at 194.

¹⁹⁰ Trial Transcript at 26, *People v. Vincent Jenkins*, No. 82-1320-001 (N.Y. Sup. Ct. June 2, 1983).

¹⁹¹ *Id.* at 27 (“In order for a comparison to be made, two strands of hair would have to be exactly identical, and in this particular case, it varied in the diameter, which is measured under a microscope. The hair taken off the victim had been cut, the end had been cut, the hair taken from Mr. Jenkins had a tapered end, which meant that it was not cut. And the hair taken from Mr. Jenkins had a different medulla, which is the center part of the hair. And, in general, I cannot possibly say that the two hairs were similar.”).

¹⁹² *Id.*

match.¹⁹³ Krajewski testified there was a statistical possibility that other hairs on Abdal's body might match by citing to a study in an invalid way:

The study shows it would not be unusual to have to look at 4,500 strands of hair from the head in order to get a match with any one particular hair. And, from the pubic hair, one may have to look at as much as 800 hairs, and it can be from the same person. That gives an idea of how much a hair can vary just within one single person.¹⁹⁴

Again, there is no empirical data for the frequency of different characteristics that hair microscopists compare. Krajewski's testimony—even if reliance on the study referred to were appropriate—would suggest a statistical basis *not* to rely on the forensic method of hair comparison, which is based on selected exemplar hairs rather than on review of hundreds of hairs from a given person.

No hair comparison, resulting in either inclusion or exclusion of an individual, could be reliable if it were true that human hairs exhibit such variation. No studies have addressed that crucial question. Nor have any analysts in these cases suggested such reasons to doubt the methodology used when they readily concluded that hairs were similar. Instead, in these exonerees' trials, hair examiners made a range of invalid claims purporting to individualize hairs based on probabilities, supposedly "unusual" characteristics, or the examiner's undocumented experience. None of these statements were supported by empirical data.

C. Invalid DNA Testimony

Three of the 11 exoneree trial transcripts obtained that had testimony concerning DNA testing contained invalid testimony concerning the DNA testing. Of the other 8 cases, 1 involved a gross error in analysis; the last 7 involved DNA that excluded the defendant at the time of the criminal trial, in 3 of which the defense

¹⁹³ Id. at 36–37.

¹⁹⁴ Id. at 37–38. That study and the consensus that it should not be relied on to suggest a statistical basis for microscopic hair comparison is discussed *supra* note 158. In addition, the testimony misrepresents the erroneous conclusions of that study.

called the only expert witness.¹⁹⁵ The first invalid DNA testimony case, that of Gilbert Alejandro, involved egregious testimony by Fred Zain, who testified that he had conducted DNA testing and obtained results inculcating Alejandro. He told the jury, “the banding patterns that were identified from these items that you mentioned were identical to the banding patterns of Mr. Alejandro. As I stated in the report, they could only have originated from him.”¹⁹⁶ He gave no random match criteria for this supposed DNA inclusion, but falsely told the jury that “DNA typing is a hundred percent identity as to whether a blood or body fluid may have originated from a particular donor or not.”¹⁹⁷ A subsequent internal inquiry concluded that Zain had at best compared only partial banding pattern results visually; later tests excluded Alejandro.¹⁹⁸

State of Texas v. Josiah Sutton

In the Josiah Sutton case, the victim had been raped by two men in the back seat of her car. Semen was present in the vaginal swab and on the stain removed from the back seat where the rape occurred. The official report authored by the Houston Police Department Crime Laboratory and the trial testimony of laboratory analyst Christy Kim presented invalid DNA results.¹⁹⁹ The raw data and the analyst’s bench notes indicate that whereas the vaginal sample reflected a mixture of the victim’s DNA and DNA from two male donors, the semen stain on the car seat suggested it came from one man only and that the lone male could not be Sutton. Although Sutton was excluded as the source of the car seat semen stain, that conclusion was not mentioned in the official report nor

¹⁹⁵ Those 7 cases are those of R. Alexander, J. Deskovic, C. Elkins, N. Hatchett, T. Hayes, E. Karage, and R. Mathews.

¹⁹⁶ Trial Transcript at 149, *State v. Gilbert Alejandro*, No. 90-09-8445-CR (Tex. Dist. Ct. Dec. 11, 1990).

¹⁹⁷ *Id.* at 146.

¹⁹⁸ See Evaluation of Out-of-County Serology Cases, Memorandum from Lonnie D. Ginsberg, Chief Forensic Serologist, Bexar County Forensic Science Center Criminal Investigation Laboratory, to Vincent DiMaio, Chief Medical Examiner (June 28, 1994); The Innocence Project, Profile: Gilbert Alejandro, <http://www.innocenceproject.org/Content/47.php> (last visited Nov. 10, 2008).

¹⁹⁹ See Trial Transcript at 168–230, *State v. Josiah Sutton*, No. 800450 (Tex. Dist. Ct. 1999), available at <http://www.scientific.org/archive/Christy%20Kim%20Testimony.pdf> [hereinafter Transcript, Sutton].

in the analyst's testimony at trial. Most importantly, if the back seat stain was considered in conjunction with the vaginal swab data, Sutton could probably have been ruled out as one of the rapists.²⁰⁰

Instead, the report erroneously concluded that the DNA profile on the seat stain was consistent with a mixture from Sutton, the victim, and another man. The report then states: "The DNA type of J. Sutton can be expected to occur in 1 out of 694,00[0] people among the black population."²⁰¹ But as explained in the section on serology, this frequency is irrelevant and misleading. The only relevant statistic is the combined frequency of all potential semen donors. Since the sample was supposedly a mixture of two male donors and it was impossible in this case to distinguish primary and secondary donors, the correct statistic for characterizing the value of the evidence is the sum of the frequencies of all possible donors. The final result would have revealed a potential donor population far larger and an event far more common than reported.²⁰²

At trial, Kim presented no statistics. However, she gave testimony that implied uniqueness for each DNA pattern and said that Mr. Sutton's DNA pattern was detected in the evidentiary samples.²⁰³ Kim testified, for example, "If it came from one person, it should have a same exact DNA pattern. No other two persons will

²⁰⁰ See William Thompson, Review of DNA Evidence in *State of Texas v. Josiah Sutton* (2003), available at <http://www.scientific.org/archive/Thompson%20Report.PDF>; see also Thompson, *supra* note 59, at 107–19. Since the victim had denied recent sex and indicated that the only event that could have produced the semen on the back seat was the rape, in all likelihood the single male profile on the stain from the seat—which excluded Sutton—was deposited by one of the two rapists. The profile from the seat stain is also consistent with one of the two male profiles contained in the vaginal swab. By a process of elimination, the genetic profile of the second rapist can be inferred. That second profile is also inconsistent with Mr. Sutton. In other words, the DNA evidence taken as a whole provides strong evidence of Sutton's innocence. The jury knew nothing about this exculpatory evidence. Cf. Michael R. Bromwich, Final Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room 210 (2007), available at <http://www.hpdlabinvestigation.org/reports/070613report.pdf> (agreeing with Thompson's conclusion that test results on Sample #1 were not properly reported, but suggesting that they should have been reported as an exclusion as to that sample, but with an inconclusive result due to "Ms. Kim performing poor DNA testing on a potentially probative sample").

²⁰¹ Thompson, *supra* note 200, at 7.

²⁰² *Id.* (calculating the frequency of possible contributors as 1 in 15, or 1 in 8 for black men in Texas).

²⁰³ Transcript, Sutton, *supra* note 200, at 181, 184–85.

have [the] same DNA except in the case of—of identical twins.”²⁰⁴ The jury was left with the mistaken impression that the DNA evidence uniquely identified Sutton as the rapist.²⁰⁵

State of Florida v. Chad Richard Heins

In Chad Heins’s case, examiners at the Florida Department of Law Enforcement conducted then-available first-generation DNA testing. When they testified, the examiners failed to report the percentage of the population that could have contributed to the mixed samples found in a sink drain plug and washcloth, and instead left the impression that Chad Heins, who they explained had a genetic marker shared by only 8.5% of the population, could have contributed to the sample.²⁰⁶ This testimony did not play a dispositive role in the case, however, because Chad Heins lived in the house from which the samples were taken, and could have been expected to have used the washcloth and sink. Ultimately, more sophisticated STR testing would show that the same unknown person’s DNA profile was found in the sink drain, washcloth, and hairs.

A final case, that of Timothy Durham, involved not faulty testimony concerning DNA analysis (though the hair comparison testimony was invalid), but rather gross error in conducting the testing and interpreting the results.²⁰⁷

²⁰⁴ *Id.* at 176.

²⁰⁵ See Thompson, *supra* note 200, at 8.

²⁰⁶ See Trial Transcript at 1158–59, *State v. Chad Richard Heins*, No. 94-3965-CF (Fla. Cir. Ct. Dec. 13, 1994). The DQ Alpha testing used only typed a single genetic marker, generating frequencies like in this case of 1 in 12. In contrast, modern STR DNA testing used today types 13 genetic markers and is capable of generating frequencies of one in many billions. See, e.g., Nat’l Comm’n on the Future of DNA Evidence, U.S. Dep’t of Justice, *The Future of DNA Testing* 19 (2000).

²⁰⁷ See William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 *J. Forensic Sci.* 47, 48 (2003) (“The initial DNA test result that helped convict Durham was proven to have been a false positive. The error arose from misinterpretation. The laboratory had failed to completely separate male from female DNA during differential extraction of the semen stain. The victim’s alleles, when combined with those of the true rapist, produced an apparent genotype that matched Durham’s. The laboratory mistook this mixed profile for a single source result, and thereby falsely incriminated an innocent man. Durham was released from prison in 1997.”).

D. Invalid Bite Mark Testimony

Forensic odontology includes two very different disciplines. The older discipline involves the identification of a decedent by matching a well-preserved and three-dimensional set of teeth to dental records. Dentists perform a vital service in distinguishing among the dead in mass disasters such as plane and train crashes. X-rays collected from the deceased's dental records can be readily matched to the fillings and crowns observed in the mouth of the deceased. The second, and certainly more controversial, application involves the interpretation of lacerations, abrasions, and bruises of questionable origin on skin and, in particular, on decomposing skin. Although the small forensic dental community refers to the discipline as "bite mark" analysis, often the most challenging conclusion is the threshold question of whether the marks are due to a human bite as opposed to some other post mortem artifact. Unlike the wax mold a dentist makes in her office, skin, given its elasticity and potential for distortion, is a poor medium for accurately registering the bite marks.

Such bite mark work is "based on the assumption that every person's dentition is unique," though this assumption has not been tested.²⁰⁸ Indeed, the NAS report noted that not only do "no scientific studies support" the assumption "that bite marks can demonstrate sufficient detail for positive identification," but that "[s]ome research is warranted in order to identify the circumstances within which the methods of forensic odontology can provide probative value."²⁰⁹ After all, even if the assumption of dental uniqueness were established as true, that uniqueness would be far easier to identify from pristine wax molds made in a dentist's office than from the few distorted impressions left on the skin during a very dynamic biting situation. Nevertheless, courts permitted forensic

²⁰⁸ 1 Giannelli & Imwinkelried, *supra* note 78, § 13-2; see also C. Michael Bowers, *The Scientific Status of Bitemark Comparisons*, in 3 *Modern Scientific Evidence* 538, 549-50 (David L. Faigman et al. eds., 2002) (reviewing "less than persuasive" literature and concluding that "[t]he demonstration of uniqueness is a blend of art and opinion"); Paul C. Giannelli, *Bite Mark Analysis*, 43 *Crim. L. Bull.* 930, 931-36 (2007).

²⁰⁹ See *Strengthening Forensic Science*, *supra* note 12, at 5-37.

odontology testimony in the 1980s and continue to do so; “no reported case has rejected bite mark evidence.”²¹⁰

While “there is no quantitative base for bitemarks analysis,”²¹¹ the guidelines promulgated by the American Board of Forensic Odontology (“ABFO”) permit its members to render conclusions expressing near certainty—they may conclude that a bite mark matches a criminal defendant to a “reasonable medical certainty” and “high degree of certainty,” explaining that the intended connotation is a “virtual certainty; no reasonable or practical possibility that someone else did it.”²¹² The guidelines counsel that, while experts may not convey “unconditional certainty,” they may express “reasonable medical certainty”; moreover, “It is . . . acceptable to state that there is ‘no doubt in my mind’ or ‘in my opinion, the suspect is the biter’ when such statements are prompted in testimony.”²¹³ No scientific criteria exist for what observations and analysis permit an expert to draw each type of conclusion.²¹⁴ Indeed, analysts conclude that variation between the bite mark and the defendant’s teeth need not disturb a finding that the bite marks are consistent, and no guidelines explain which points of comparison are required for a positive identification (an ABFO effort in 1984 to adopt a scoring system was abandoned).²¹⁵

²¹⁰ Paul C. Giannelli, *Bite Mark Evidence*, GP Solo (Sept. 2007), available at <http://www.abanet.org/genpractice/magazine/2007/sep/trialpractice-bitemark.html>; see also Giannelli, *supra* note 208, at 933 n.15 (“The overall ‘uniqueness’ of dental characteristics is a common statement used in court and in literature. This conclusion is generally accepted but is subject to considerable criticism. The reason it is criticized is that it has never been proven.” (quoting C. Michael Bowers, *Forensic Dental Evidence: An Investigator’s Handbook* 197 (2004))).

²¹¹ Iain A. Pretty, *Reliability of Bitemark Evidence*, in *Bitemark Evidence* 531, 543 (Robert B.J. Dorion ed., 2005).

²¹² ABFO Bitemark Methodology Guidelines, <http://www.abfo.org/bitemark.htm> (last visited Nov. 10, 2008).

²¹³ *Id.*

²¹⁴ See Bowers, *supra* note 208, at 565 (“The center point of disagreement amongst odontologists is the issue, ‘what is necessary to support a positive identification from a bitemark?’ The odontological literature is silent on the sufficiency of evidence necessary to accomplish this task, yet this positive opinion is permitted to any dentist.”).

²¹⁵ See 1 Giannelli & Imwinkelried, *supra* note 78, § 13-2, -4. A 1999 ABFO study involving matching of four bite marks to seven sets of teeth produced a 63.5% false positive rate, and found “less than optimal accuracy.” Kristopher L. Arheart & Iain A. Pretty, *Results of the 4th ABFO Bitemark Workshop—1999*, 124 *Forensic Sci. Int’l* 104 (2001) (noting that the study used only medium-to-good forensic quality bite marks); Bowers, *supra* note 208, at 545 (calculating false positive rate not presented in

Six cases involved bite mark comparison, and trial transcripts were located for all 6: the cases of Kennedy Brewer, Roy Brown, Ray Krone, Willie Jackson, James O'Donnell, and Calvin Washington. In all but James O'Donnell's and Calvin Washington's cases (in which the odontologist merely observed a consistency), the odontologists provided invalid testimony. In the Brewer, Krone, and Jackson cases, the odontologists testified they were certain that the defendant left the bite marks.

People of the State of New York v. Roy Brown

In Roy Brown's case, the defense presented an expert, Homer Campbell, who concluded that the bite marks were inconsistent with Brown's teeth because, among other manifest differences, one showed impressions of six teeth from the upper bite where Roy Brown had only four.²¹⁶ The prosecution never disclosed to the defense that the State Police forensic dentist, Lowell A. Levine, had previously opined to the prosecutor that the bite marks "excluded" Brown.²¹⁷ Instead, the prosecutor presented the testimony of Edward Mofson, who found the bite marks similar to "a reasonable degree of dental certainty" and called the differences "[i]nconsistent but explainably so in [his] opinion."²¹⁸

State of Arizona v. Ray Milton Krone

Ray Krone's case was particularly troubling, for the bite mark evidence played a "critical" role: two experts concluded that the defendant made the bite mark on the victim, and Krone was then sentenced to death.²¹⁹ Experienced forensic odontologist Raymond Rawson presented the bite mark evidence at trial, along with John Piakis, a dentist who was inexperienced and just beginning to serve

the ABFO results and noting, "If this reflects their performance in actual cases, then inculpatory opinions by forensic dentists are more likely to be wrong than right").

²¹⁶ Trial Transcript at 951, 953, *People v. Roy Brown*, No. 91-2099 (N.Y. County Ct. Jan. 13-23, 1992) [hereinafter Transcript, R. Brown].

²¹⁷ Fernanda Santos, *With DNA From Exhumed Body, Man Finally Wins Freedom*, N.Y. Times, Jan. 24, 2007, at B5.

²¹⁸ Trial Transcript, R. Brown, *supra* note 216, at 740, 774.

²¹⁹ *State v. Krone*, 182 Ariz. 319, 322 (1995) ("The bite marks on the victim were critical to the State's case. Without them, there likely would have been no jury submissible case against Krone.").

as the police department's odontologist. Rawson presented a highly inflammatory and unusual video with images of Piakis holding molds of Krone's teeth to the marks on the deceased victim's body. Rawson attempted to quantify the probability of a tooth association:

And it turns out that on average a tooth can be in about 150 different positions, each one of which is easily recognizable. And if you are looking at a tooth in that kind of detail, then you can see that very quickly. Just having two teeth, the possibilities of two teeth being in the same position, it would be 150 times 150, whatever that is. Maybe 1200 or something like that.²²⁰

Rawson then told the jury in no uncertain terms that Krone had left the bite marks:

A. That's as nice a match as we—as we really ever see in a bite mark case.

Q. By “nice” do you mean accurate?

A. Yes. That was a nonscientific term. This is really an excellent match, and would be held in high regard by forensic odontologists.

Now there's a wiping action just to show the same thing. Again, high correlation. I mean, that is—that tooth caused that injury.²²¹

He concluded his testimony agreeing that “it was Ray Krone's teeth.”²²² Piakis similarly testified, “I say that there is a match. Okay? I'm saying there's a definite match.”²²³

The defense never learned that, before trial, police had initially consulted FBI odontologist Skip Sperber, who after examining the bite marks concluded, “It could not have been clearer. . . . Ray Krone had two higher teeth than his incisors that would have marked when he bit. Those weren't there in the evidence.”²²⁴ The

²²⁰ Trial Transcript at 15, *State v. Ray Milton Krone*, No. CR 92-00212 (Ariz. Super. Ct. Aug. 4, 1992).

²²¹ *Id.* at 39.

²²² *Id.* at 57.

²²³ *Id.* at 91.

²²⁴ Robert Nelson, *About Face*, *Phoenix New Times*, Apr. 21, 2005, <http://www.phoenixnewtimes.com/2005-04-21/news/about-face/>.

2009]

Invalid Forensic Science Testimony

71

Chicago Tribune later reported, “The discrediting of Rawson’s testimony in the Krone case is one of numerous instances in which leading practitioners of bite-mark comparison have erred.”²²⁵

*E. Additional Forensic Techniques**1. Shoe Print Analysis*

Three cases in the study set involved shoe print testimony. Two of the cases involved shoe prints that either excluded the defendant or were non-probative. The third is the case of Charles Fain.

State of Idaho v. Charles I. Fain

In Charles Fain’s case, Robert B. Hallett of the FBI testified that the make of the shoe print was consistent with Fain’s, and that “[i]t was possible that this shoe made this impression.”²²⁶ Not satisfied with his initial cautious conclusion, Hallett added that, although it was a common type of boat shoe sole, the wear patterns on the shoe individualized the print:

Q. Okay, you also, if I understand correctly, that you said if another shoe made the impression, it would have to have the same characteristics as the actual left shoe that we have here?

A. That’s correct, sir.

Q. What are those characteristics?

A. The same size, the same design, and having general wear in exactly the same locations.

Q. Now, did you indicate that the wear characteristics are put there by a gait of a particular individual?

A. You would have to have the same characteristic walk as the individual who owned those shoes.

Indeed, Hallett also testified so as to imply that other examiners might have reached even stronger conclusions:

²²⁵ Flynn McRoberts, *Bite-Mark Verdict Faces New Scrutiny*, *Chi. Trib.*, Nov. 29, 2004, at 1.

²²⁶ Trial Transcript at 281, 294, *State v. Charles I. Fain*, No. C-5448 (Idaho Dist. Ct. Oct. 14, 1983).

I found, therefore, that the shoe which made this impression, and this left shoe had sustained wear in the same area. To a shoe print examiner, this would indicate that the individual who walked with these shoes has the same walking gait.

Some examiners believe, I have not quite gone that far myself, but that could be a positive identifying characteristic. They believe we all walk differently.

That wear corresponded exactly.²²⁷

This practitioner suggested that the effect of gait on the sole of a shoe is unique. No data supports such an opinion. Adding a clever embellishment, he testifies that other examiners would go even further to say that wear patterns on shoes “correspond[] exactly.” Unfortunately, that is the case: other examiners may indeed go further in their testimony, on the recommendation of the Scientific Working Group on Shoeprint and Tire Tread Evidence (“SWGTTREAD”), which offers the guideline that an examiner can find an “[i]dentification (definite conclusion of identity).”²²⁸ The guideline explains that “this is the highest degree of association expressed in footwear and tire impression examinations. This opinion means that the particular shoe or tire made the impression to the exclusion of all other shoes or tires.”²²⁹ No scientific criteria are provided regarding when an expert may render any of those conclusions.²³⁰

2. Fingerprint Analysis

Fingerprint comparisons were conducted in 14 exonerees’ cases.²³¹ Trial transcripts were located for 13 of these cases. Two in-

²²⁷ Id. at 298.

²²⁸ Scientific Working Group on Shoeprint and Tire Tread Evidence, Int’l Ass’n for Identification, Standard Terminology for Expressing Conclusions of Forensic Footwear and Tire Impression Examinations 1 (Mar. 2006), http://www.theiai.org/guidelines/swgtread/terminology_final.pdf.

²²⁹ Id.

²³⁰ See Yaron Shor & Sarena Weisner, A Survey on the Conclusions Drawn on the Same Footwear Marks Obtained in Actual Cases by Several Experts Throughout the World, 44 J. Forensic Sci. 380, 383 (1999) (finding a wide range of variability in criteria experts use to draw conclusions in shoe print cases).

²³¹ In several additional cases, examiners did not conduct comparisons because latent prints were unsuitable for comparison.

involved troubling testimony or analysis; the others all involved fingerprint exclusions at trial (and in one the only expert testified for the defense).²³² In the first troubling case, that of Gene Bibbins, as discussed further below, the examiner testified that the comparison between his fingerprints and latent prints found on the window fan in the victim's room was non-probative, when in fact the Louisiana State Crime Lab had excluded Bibbins and documented its contrary finding in a report not disclosed to the defense.²³³

Commonwealth of Massachusetts v. Stephan Cowans

In the case of Stephan Cowans, a Boston police officer was shot by a civilian. In the course of the assailant's escape, the assailant picked up a glass mug, drank from it, and put it back down. The crime scene unit promptly vouchered the mug and lifted two latent prints from it. After Cowans became a suspect, Boston Police latent expert Dennis LeBlanc compared Cowans's known ink thumb print to one of the latent prints and declared a match. The second print was favorably compared to the woman who owned the mug. After Cowans was exonerated by post-conviction DNA testing, the District Attorney asked the Massachusetts State Police to reexamine the thumb print. The State Police declared that Cowans was clearly excluded.²³⁴

The Boston Police then hired an external auditor, Ron Smith & Associates, to conduct an independent investigation into, among other things, Officer LeBlanc's conduct in the case. The audit team had four members, all experts in fingerprint comparison.²³⁵ The auditors reached the unanimous conclusion that Officer LeBlanc realized at some point prior to trial that Cowans was excluded, but

²³² The other cases were those of Antonio Beaver, Michael Cromedy, Frederick Daye, James Giles, Dennis Halstead, Anthony Hicks, Larry Mayes, John Restivo, Samuel Scott, James Tillman, and Ron Williamson. Cromedy, Daye, Giles, Halstead, Hicks, Mayes, Restivo, and Tillman's cases involved fingerprint exclusions. Scott's case involved a conclusion that a fingerprint belonged to Scott, a conclusion that was not terribly probative since it was found on his glass in his house. Curtis McCarty's case also involved a matched latent print; that portion of his trial transcript, however, has not been located.

²³³ See discussion *infra* Subsection II.F.1.

²³⁴ See Possley, *supra* note 9.

²³⁵ Ron Smith & Associates, Inc., Reference: Request for Latent Print Consultation Services 6 (Mar. 8, 2004).

that he nevertheless concealed that fact in his trial testimony.²³⁶ Instead, Officer LeBlanc misrepresented to the jury that the latent print matched Cowans's. The auditors' conclusion was based on facts including: Cowans's exclusion was clear to every member of the review team; Officer LeBlanc had made correct associations and exclusions routinely in more difficult cases over the preceding four years; he made efforts to conceal other errors made in the same case; there were numerous inconsistencies in his testimony; and he intentionally used a method of showcasing the erroneous Cowans match evidence to the jury that not only made it more difficult for the jury to follow but was contrary to the preferred methods of fingerprint examiners and contrary to what Officer LeBlanc did with the other latent print in the same case.²³⁷

In the other fingerprint cases, the evidence played little role. For example, in the cases of James Giles and Michael Cromedy, the examiners testified that the prints excluded the defendants; similarly, in the Dennis Halstead and John Restivo cases, the fingerprints did not match any known person.

3. Voice Analysis

One final case, that of David Shawn Pope, involved voice comparison using a spectrograph, an instrument that generates a visual pattern depicting an audio recording using lines that represent the frequency and intensity of the sound wave over time.²³⁸ Although the National Academy of Sciences ("NAS") issued a report in 1979 concluding that the use of voice spectrograph analysis to identify individuals "is not adequately supported by scientific data" (after which the FBI stopped permitting court testimony concerning such analysis), a series of courts continue to permit testimony concerning voice spectrography, apparently acting "as if the NAS Report did not exist."²³⁹

In Pope's case, the victim of a 1985 rape in Garland, Texas, received several messages on her answering machine shortly after the crime. The Dallas County police arrested Pope after the victim

²³⁶ Id. at 26–27.

²³⁷ Id.

²³⁸ See The Scientific Basis of Expert Testimony on Talker Identification, in 5 Modern Scientific Evidence § 37-1 (David L. Faigman et al. eds., 2007).

²³⁹ Id.

2009]

Invalid Forensic Science Testimony

75

identified him in a lineup, and later conducted voice spectrograph analysis comparing a recording of Pope's voice with the messages on the victim's answering machine. The State retained Henry Truby as an expert. He testified, based on finding "10–15 similar patterns" shared by the recording of Pope and the recording from the victim's answering machine, that "the original producer of [the recordings] was the same individual."²⁴⁰ Truby explained:

A. I found a sufficient number [of identical patterns] to serve as an identification to convince me, and then take a few more just to reinforce it, that no matter how much you do of these samples, you would continue to get points of similarity every now and then.

Q. All right. Let me ask you then, so that it is clear, are you saying the known tape and the unknown tape were made by one and the same person?

A. I do so state.

Truby testified at the end of his direct examination:

Q. The bottom line analysis on the known voice and the unknown voice in this situation were only made by one single person in the whole wide world?

A. Exactly.

Q. Just like fingerprints, it is unique?

A. Exactly.²⁴¹

In addition to voice spectrography being found unreliable by the NAS panel and barred from use in court by the FBI, no study has suggested that an analyst can conclude that only one person in the world could produce a particular pattern exhibiting certain similarities with an unknown pattern. Indeed, the defense retained Stuart I. Ritterman, a professor of communicology at the University of South Florida, who testified that studies show that spectrography

²⁴⁰ Trial Transcript at 290, *State v. David Shawn Pope*, No. F85-98755-NQ (Tex. Dist. Ct. Feb. 4, 1986).

²⁴¹ *Id.* at 295.

“is totally unsuitable as a tool for indentifying voices with any degree of accuracy.”²⁴²

F. Forensic Science Misconduct Beyond Trial Testimony

1. Withholding Forensic Evidence

Thirteen cases were identified as involving either a failure to disclose exculpatory data or analysis, or outright fabrication of evidence.²⁴³ Examples included withholding laboratory reports, analysis, or the existence of evidence. Other cases involved fabrication, including falsifying or altering lab reports. Putting aside the examples of fabrication, this study does not opine whether evidence was withheld due to deliberate, reckless, or negligent acts.²⁴⁴ The known failures to disclose forensic evidence helpful to the defense remain only a subset; other evidence of innocence may remain undisclosed to this day. From the trial transcripts alone it is impossible to know whether material was concealed. Even with the benefit of bench notes or laboratory reports, one may not be able to ascertain whether experts falsified or concealed test results.

Commonwealth of Virginia v. Earl Washington, Jr.

Data or analysis was withheld in cases involving a number of forensic disciplines. Several cases involved serology. One example is in the case of Earl Washington, a man sentenced to death in Virginia for a rape and murder that he did not commit. The victim was a Caucasian woman who stated before she died from severe knife wounds that the attacker was African-American. The state never disclosed to the defense a serology report, conducted early in the investigation and dated August 19, 1982, which found that stains on a central piece of evidence, a light blue baby's blanket on the murdered victim's bed, were not only ABO Type A, PGM Type 1, but

²⁴² Id. at 321, 329.

²⁴³ The cases are those of G. Bibbins, R. Brown, S. Cowans, W. Gregory, R. Krone, C. McCarty, N. Miller, J. Sutton, E. Washington, and K. Waters.

²⁴⁴ State of mind is not relevant to the inquiry under *Brady v. Maryland*, 373 U.S. 83, 87 (1963). In these cases the evidence did not surface until after post-conviction DNA testing, post-exoneration investigations, or civil suits.

also “Tf CD.”²⁴⁵ As a subsequent October 22, 1982, police report noted, Transferrin CD is a fairly uncommon plasma protein that is most often found in African-Americans. Virginia Bureau of Forensic Science examiner Deanne Dabbs told the Virginia State Police that “the Tf CD is found in 10% of Negroes but to her knowledge has never been found in Caucasians. In order to be sure of this determination, she stated she ran a second test with the same results.”²⁴⁶

This highly probative serology finding regarding the Tf CD was never disclosed to the defense. Instead, the initial report was altered, but only after Earl Washington was identified as a suspect in 1983. Earl Washington did not possess the unusual “Tf CD” characteristic. Rather than report to the defense that Washington lacked the Tf CD that had been observed on the stains, the state created a second “amended” report, dated August 23, 1983, and provided it to the defense. That second report, issued without having conducted any further testimony undermining the original results, nevertheless stated that “[t]he results of Tf typing were inconclusive.”²⁴⁷ Thus, law enforcement concealed probative information regarding the blood type of the perpetrator.

Other cases involved concealment of exculpatory information regarding hair comparison. In William Gregory’s case, Dawn Katz concealed the fact that she determined that at least one hair was not consistent with Gregory’s hair.²⁴⁸ Joyce Gilchrist concealed and altered laboratory reports in which she had initially excluded Curtis McCarty, which led a court to grant a new trial because of her *Brady* violations.²⁴⁹

²⁴⁵ Certificate of Analysis, Commonwealth of Virginia Bureau of Forensic Science (Aug. 19, 1982).

²⁴⁶ SA C.R. Wilmore, Virginia State Police Report (Oct. 22, 1982) (emphasis omitted).

²⁴⁷ Amended Copy of Certificate of Analysis dated August 19, 1982, Commonwealth of Virginia Bureau of Forensic Science (Aug. 26, 1983).

²⁴⁸ See *Gregory v. City of Louisville*, 444 F.3d 725, 732 (6th Cir. 2006) (affirming denial of absolute or qualified immunity to Katz, stating that “Katz’s deposition for this instant action revealed that Katz had actually found 7 negroid head hairs on the pantyhose, only 5 of which she found similar to Plaintiff’s hair”).

²⁴⁹ *McCarty v. State*, 114 P.3d 1089, 1092, 1095 (Okla. Crim. App. 2005); see also *The Innocence Project, Curtis McCarty*, <http://www.innocenceproject.org/Content/576.php> (last visited Sept. 14, 2008).

In two bite mark comparison cases already noted—those of Roy Brown and Ray Krone—the state concealed that bite marks had been shown to other odontologists who excluded the defendant. The Josiah Sutton case discussed earlier involved gross overstatement of DNA results and a “failure to report aspects of the DNA evidence that appear[ed] to exonerate Josiah Sutton.”²⁵⁰ Two fingerprint cases, including the Stephan Cowans case described above,²⁵¹ involved the withholding of exculpatory evidence.

State of Louisiana v. Gene Bibbins

In Gene Bibbins’s case, Annie Michelli of the Baton Rouge City Police had testified at trial that any comparison between Bibbins’s fingerprints and a latent print found on the window at the crime scene was inconclusive, explaining that “[t]he latents are unidentifiable. You can’t—they aren’t any—there aren’t any prints on there that we can use.”²⁵² When asked, “Did you double-check your conclusion with the state crime lab?” and, “Did they have the same results?” she answered to both questions, “Yes, Ma’am.”²⁵³ That testimony was false; the state crime lab’s finding and report had excluded Bibbins. The district court in a civil rights lawsuit filed by Bibbins denied Michelli’s motion for summary judgment on Bibbins’s *Brady* claim, stating:

Michelli’s testimony at Bibbins’[s] trial was that she was unable to identify Bibbins as a match to the fingerprint sample. Michelli also testified that she double checked her results with the Louisiana state crime lab and that the state crime lab reached the same results. However, it is undisputed that a Louisiana state police crime lab report made by Sybil Guidry showed a contrary result. Guidry’s findings excluded Bibbins as a match.²⁵⁴

Additional cases involving the withholding of exculpatory evidence have been documented, but the authors have not yet obtained transcripts; thus, those cases were not included in this

²⁵⁰ See Thompson, *supra* note 200, at 2.

²⁵¹ See *supra* Section II.E.

²⁵² Trial Transcript at 83, *State v. Gene C. Bibbins*, No. 2-87-979 (La. Dist. Ct. Mar. 25, 1987).

²⁵³ *Id.* at 83, 84.

²⁵⁴ *Bibbins v. City of Baton Rouge*, 489 F. Supp. 2d 562, 572 (M.D. La. 2007).

study.²⁵⁵ The case of Armand Villasana also demonstrates the importance of the discovery of underlying laboratory notes. In that case, Villasana, another DNA exoneree (not included here because the trial testimony was never transcribed) later brought a civil suit because the forensic analyst never provided laboratory notes showing that material existed with which to perform DNA testing at the time of trial.²⁵⁶

2. *Gross Error in Analysis*

In a few cases, reanalysis of the evidence conducted post-conviction uncovered gross errors that were not apparent at the time of trial. Again, these cases represent only those in which errors came to light due to the rare use of post-conviction reevaluation or retesting. Few exonerees have had the forensic analysis evidence in their cases retested or reexamined. The gross errors that were uncovered include mistyping serological evidence, failing to notice abundant spermatozoa, erroneously linking large numbers of hairs, failing to note material differences during comparisons, and failing to use equipment properly. It is not known how many other exonerees, much less non-exonerated convicts, were convicted based on grossly erroneous forensic testing or analysis.

For example, in several serology cases, evidence was reported as non-probative and not more elaborately tested due to a supposed failure to observe spermatozoa; subsequent examiners who thereafter reviewed the same evidence found abundant spermatozoa. In

²⁵⁵ In a series of cases, forensic expert Pamela Fish notoriously concealed evidence. None of those transcripts were included in the above figures because they have not yet been obtained—though the transcripts would shed little light on the matter, as the issue is precisely that exculpatory evidence was not disclosed to the defendants' trial counsel and did not arise at trial. For example, in John Willis's case, Fish testified that her tests were "inconclusive," but when DNA testing exonerated him years later, her lab notes surfaced showing Willis was a Type A secretor, whereas the material tested indicated a Type B semen donor. See Paul C. Giannelli, *Bench Notes & Lab Reports*, 22 *Crim. Just.* 50, 50 (2007).

²⁵⁶ See *Villasana v. Wilhoit*, 368 F.3d 976, 979 (8th Cir. 2004) (holding forensic technician's notes underlying disclosed lab reports on DNA testing were not exculpatory, even though they led the defense to perform additional testing, because *Brady* applies only to evidence a reasonable prosecutor would identify at the time as material); cf. ABA Criminal Justice Standards for DNA Evidence Standard 16-4.1 (3d ed. 2007) (recommending that not only laboratory reports but underlying laboratory case file and notes be maintained and disclosed to the defense).

the Larry Peterson case, “[a]lthough the New Jersey State Police Laboratory had reported that there was no semen in the victim’s rape kit,” the Serological Research Institute, before conducting its post-conviction DNA testing, “identified sperm on her oral, vaginal, and anal swabs.”²⁵⁷

The Ulysses Charles case provides another example of this type of error in conducting presumptive testing. Charles was a Type B secretor, while the victims were Type O secretors and the stains on a robe and sheet contained only H antigens consistent with that O type.²⁵⁸ The prosecution called two experts to explain why no B antigens consistent with Charles were observed. Stanley Bogdan of the Boston Police Department Crime Laboratory explained that though acid phosphatase was detected through assays, P-30 testing did not indicate the presence of sperm. A second expert, John Cope Abbott, also explained the P-30 test results and noted that no spermatozoa were observed.²⁵⁹ We now know this was gross error, for when Cellmark later performed DNA testing, spermatozoa were readily observed on the same robe and sheet.²⁶⁰ The technology for identifying the presence of sperm, a conventional microscope, has remained unchanged for decades. There was no technological reason why the spermatozoa could not have been observed in 1984 when Charles was tried.

The serology cases involving conjectural theories of contamination all involve gross error. As described above, in the case of Gary Dotson, Edward Blake found gross error not only because the phenomenon of masking was not explained, but also because the analyst did not attempt to use control testing to assess whether contamination was a proper explanation for the finding of a blood type that was inconsistent with both Dotson and the victim.²⁶¹ Other serology cases involved mistyping. For example, in the Ford Heights Four case, Chicago Police Department examiner Michael

²⁵⁷ The Innocence Project, Larry Peterson, <http://www.innocenceproject.org/Content/148.php> (last visited Nov. 10, 2008).

²⁵⁸ Trial Transcript at 7-28 to -32, *Commonwealth v. Rodriguez U. Charles*, No. 035492-45 (Mass. Super. Ct. Feb. 9, 1984).

²⁵⁹ *Id.* at 7-49 to -50.

²⁶⁰ Letter from Gina Pineda, Reliagene Technologies, Inc., to Myeti Gametchu (Sept. 26, 2007) (“Cellmark microscopically observed sperm cells in each of the sperm f[r]action samples . . .”).

²⁶¹ Affidavit of Edward T. Blake, *supra* note 3 at 23-25.

Podlecki reported that Dennis Williams was a Type A secretor,²⁶² when Edward Blake found post-conviction that in fact he was a Type A non-secretor.²⁶³

Gross error in several bite mark comparison cases has just been discussed, as well as in the DNA testing conducted in the Timothy Durham case. In several hair comparison cases, reexamination of the evidence was conducted as part of a post-exoneration investigation, and reports found the comparison at the time of trial to have been erroneous. For example, in the Jimmy Ray Bromgard case discussed earlier, the FBI's reexamination revealed that the head and pubic hairs that Arnold Melnikoff had found microscopically indistinguishable were in fact "microscopically dissimilar" to Bromgard's, and that the head hair was in fact similar to the victim's.²⁶⁴

Additional examples in which hair evidence matched the actual perpetrator are discussed in the next Section. Additional egregious examples of error in hair comparison cases include cases where the examiner compared large numbers of hairs and still erroneously linked all of them to an innocent man. As the Bromgard Peer Review Report concluded, "While an experienced hair examiner might erroneously associate a single head or pubic hair, it is highly unlikely that a competent hair examiner would incorrectly associate both head and pubic hairs."²⁶⁵ In quite a few cases, many more than one or two hairs were incorrectly associated with the defendant. For example, in the Curtis McCarty case, Joyce Gilchrist not only altered lab reports, but found dozens of hairs to have been consistent with McCarty.

3. Failures to Conduct Elimination Testing or Comparison

Related to the problem of gross error, forensic analysts in several cases stated that they failed to conduct serology testing on relevant potential sources. For instance, if the semen excludes the

²⁶² See Transcript, Rainge, *supra* note 168, at 2281.

²⁶³ See Center for Wrongful Convictions, Dennis Williams, <http://www.law.northwestern.edu/wrongfulconvictions/exonerations/ilWilliamsChart.pdf> (last visited Dec. 1, 2008).

²⁶⁴ See Adam Liptak, 2 States to Review Lab Work of Expert Who Erred on ID, *N.Y. Times*, Dec. 19, 2002, at A24.

²⁶⁵ See Peer Review Report, *supra* note 159, at 2.

defendant, one would want to perform testing on a potential consensual donor such as a husband or boyfriend. However, neither prosecutors nor analysts have a legal duty to pursue exculpatory evidence; they need only disclose exculpatory evidence that they uncover.

Commonwealth of Massachusetts v. Neil Miller

In the Neil Miller case, a semen stain on the sheet where the victim was raped produced both B and H blood group substances, but neither the victim nor Miller possessed the B antigen. At trial, the prosecution implied in the opening arguments that the Type B semen came from the boyfriend of the victim's roommate: "A sample of this defendant's blood and saliva was later obtained, and it was determined that the semen from the sheet was not the defendant's semen. [The victim] will testify that her roommate did have a boyfriend who sometimes stayed overnight."²⁶⁶ Although this hypothesis could easily have been tested, it never was. Post-conviction DNA testing revealed that the semen stain on the sheet had in fact been deposited by the rapist—not by Miller nor by the boyfriend.²⁶⁷

State of Oklahoma v. Ronald Keith Williamson

Several exonerees' cases involved not only false positives, but also false negatives. In several cases, elimination testing was not done on a known suspect that subsequent DNA testing revealed to have been the actual perpetrator. For example, in Ronald Williamson's case, Melvin R. Hett, a supervisor at the Oklahoma State Bureau of Investigation, Northwest Regional Laboratory, testified unequivocally that he had compared the hairs of the state's star witness, Glen Gore, with those at the crime scene. He testified, "I did, direct comparison with the unknown hairs," and when asked if any of Gore's hairs were microscopically consistent with the questioned hairs, he testified, "No, sir."²⁶⁸ Later, during Williamson's appeal, the Court of Appeals for the Tenth Circuit reviewed Hett's

²⁶⁶ Transcript, N. Miller, *supra* note 64, at 1-169.

²⁶⁷ See Frontline, *Burden of Innocence*, <http://www.pbs.org/wgbh/pages/frontline/shows/burden/profiles/miller.html> (last visited Nov. 10, 2008).

²⁶⁸ Trial Transcript at 736-37, *State v. Ronald Keith Williamson*, No. CRF 87-90 (Okla. Dist. Ct. Apr. 21-28, 1988).

lab report, which revealed that Hett's testimony was false. "In fact, the hair expert compared samples from Mr. Gore with hairs already determined to be consistent with those of the victim, [co-defendant] Mr. Fritz, and Mr. Williamson, but [did] not compare Mr. Gore's samples with unidentified hairs."²⁶⁹ Indeed, Hett also opined on the significance of a "match" of 17 hairs, including both scalp and public hairs. It was later determined that none of the hairs belonged to Ronald Williamson or his co-defendant Dennis Fritz. Glen Gore—whose hair exemplars were, contrary to Hett's false testimony, never compared to the "unidentified hairs"—was shown by post-conviction DNA testing to have been the actual perpetrator.²⁷⁰

State of Oklahoma v. Robert Miller

In the Robert Miller case, Joyce Gilchrist excluded as the originator of the questioned hairs a suspect who was later identified by post-conviction DNA testing and indicted.²⁷¹ Post-conviction analysis by Microtrace strongly disagreed with Gilchrist's findings, finding no similarities and highly varied reference samples.

These failures to rule out other possible suspects or assess whether material could have come from a partner and not the perpetrator may have occurred more often than in the known cases. Rarely did the forensic expert mention during the trial testimony whether elimination analysis was conducted. Also troubling is the fact that some experts made clear during their trial testimony that they only performed testing as requested by police or prosecutors, rather than on all materials that could be probative in the case. Failures to conduct testing occurred even as to highly probative materials and tests. For example, at least 7 exonerees were tried at a time when DNA testing was technologically available but not used.²⁷²

²⁶⁹ Williamson v. Ward, 110 F.3d 1508, 1522 n.15 (10th Cir. 1997).

²⁷⁰ See Barry Scheck et al., *Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted* 165 (2000).

²⁷¹ The Innocence Project, Robert Miller, <http://www.innocenceproject.org/Content/219.php> (last visited Nov. 10, 2008) ("The 1996 DNA tests ultimately led to Miller's exoneration and release in 1998. The tests also identified the true perpetrator, Ronald Lott, a man whose samples had been included in all rounds of testing.").

²⁷² For a discussion of those cases, see Garrett, *supra* note 55.

To conclude this Part, invalid science testimony in these exonerations cases raised a related set of problems. Analysts failed to adhere to scientific standards and instead exaggerated the probative value of evidence to make it appear more inculpatory than it in fact was. Aside from serology and DNA testing, the relevant disciplines lacked scientific standards defining the permissible scope of conclusions reached regarding comparison. Invalid testimony should come as no surprise given the lack of such standards.

III. REFORMING FORENSIC SCIENCE

This Part steps back to look at the roles of criminal justice actors and the possibilities for systemic reform originating from the scientific community. The first Section examines the existing regulation of forensic science, beginning with the roles of other criminal justice actors in these 137 trials, specifically: (1) prosecutors, who often misrepresented forensic evidence during closing arguments; (2) defense attorneys, who rarely received their own experts and rarely effectively cross-examined forensic analysts concerning invalid science testimony; and (3) judges, who when called on to rule regarding invalid forensic science testimony at trial or post-conviction rarely provided relief. Where each of those criminal justice actors failed to correct these miscarriages of justice, this Article concludes by developing a framework for national regulation and oversight of the forensic sciences. The renewed scrutiny of forensic science error may finally provide the impetus for federal legislation to ensure a sound scientific foundation for forensic sciences, including the disciplines that are the focus of this study.

A. The Roles of Judges, Prosecutors, and Defense Attorneys

1. Invalid Prosecution Use of Forensic Science

Although the trial transcripts cannot tell us whether prosecutors sought out invalid forensic testimony, they certainly display prosecutors, knowingly or not, developing and relying on such testimony.²⁷³ The Josiah Sutton case provides an example where, ac-

²⁷³ See William C. Thompson, A Sociological Perspective on the Science of Forensic DNA Testing, 30 U.C. Davis L. Rev. 1113, 1115 (1997); William C. Thompson & Edward L. Schumann, Interpretation of Statistical Evidence in Criminal Trials: The

ording to William Thompson's investigation, the prosecutor diverted the forensic expert's testimony from the fact that there was a second unknown DNA profile, because the presence of two profiles meant that Sutton was excluded—not included as the examiner had testified.²⁷⁴

Opening arguments by prosecutors, when they included descriptions of the forensic evidence that overstated its probative value, may indicate that the prosecutor had met with the forensic expert and knew about the claims being advanced (of course, whether the prosecutor knew the claims were invalid cannot be ascertained). For example, in Jimmy Ray Bromgard's case, Deputy County Attorney David W. Hoefer anticipated Arnold Melnikoff's invalid testimony, telling the jury in his opening that

the experts at the State Lab out of Missoula will come and testify, and they will tell you that that hair has the same range of microscopic characteristics as that of the defendant, and they will tell you the percentage of the population that would have that kind of hair, first for the head hair, secondly for the pubic hair, and then for the two combined.²⁷⁵

In a number of cases, the prosecutor exaggerated the testimony of the forensic analyst in closings, making claims that the forensic scientist clearly did not make. Twelve were in cases with valid testimony by all forensic analysts (an additional 6 cases included both invalid and valid analyses extended by the prosecutor). Convictions should not necessarily have been reversed on those grounds, nor did these prosecutors necessarily engage in misconduct or ethical lapses. The ethical and criminal procedure rules regarding closing statements “offer[] few standards for proper prosecutorial argument,” and though prosecutors may not misrepresent facts in evidence, they may make arguments concerning inferences to be drawn from the facts.²⁷⁶ There may be a fine line between properly

Prosecutor's Fallacy and the Defense Attorney's Fallacy, 11 *Law & Hum. Behav.* 167 (1987).

²⁷⁴ Thompson, *supra* note 200, at 1; Thompson, *supra* note 59, at 119–21.

²⁷⁵ Transcript, Bromgard, *supra* note 160, at 18.

²⁷⁶ Fred C. Zacharias, Structuring the Ethics of Prosecutorial Trial Practice: Can Prosecutors Do Justice?, 44 *Vand. L. Rev.* 45, 96 (1991); see also Ann. Model Rules of Prof'l Conduct R. 3.4(e) (6th ed. 2007) (providing that a lawyer shall not “in trial,

drawing inferences and misstating facts. Even if prosecutors draw invalid inferences or make false statements to the jury, they may not face any consequences. In reviewing such claims of prosecutorial misconduct, courts examine the severity of the alleged misconduct, the strength of the State's case, and whether curative measures were taken. In doing so, appellate courts often find any error to be harmless. Federal courts limit relief to egregious cases in which the conduct "so infected the trial with unfairness as to make the resulting conviction a denial of due process."²⁷⁷

This Article thus does not address whether any particular statements constituted misconduct, but rather emphasizes that a series of closing statements did not comport with the science and left the jury with a faulty impression of what the forensic science had actually shown. In the case of Stephen Linscott, the Illinois courts reversed the conviction based on a finding of egregious prosecutorial misconduct concerning the forensic evidence as presented during closing arguments. The Appellate Court of Illinois explained, as to the serology:

No one testified that "[the victim] was raped by a non-secretor" or that the seminal material "came from a non-secretor." The prosecutor simply made-up that piece of "evidence." The made-up evidence was doubly devastating because not only was it false, but it reduced the pool of possible assailants from a substantial percentage of the male population, or even from the entire population, to just the males in twenty percent of the population.²⁷⁸

In affirming the vacatur, the Illinois Supreme Court noted: "A prosecutor must confine his arguments to the evidence and to 'reasonable inferences' that follow from it. We believe that the prosecutor in the instant case contravened this fundamental rule."²⁷⁹ However, illustrating the difficulty of prevailing on such claims, the Illinois Supreme Court, though reversing based on statements con-

allude to any matter that the lawyer does not reasonably believe is relevant or that will not be supported by admissible evidence").

²⁷⁷ *Darden v. Wainwright*, 477 U.S. 168, 181 (1986); see also, e.g., *State v. Graves*, 668 N.W.2d 860, 877 (Iowa 2003).

²⁷⁸ *People v. Linscott*, 511 N.E.2d 1303, 1306 (Ill. App. Ct. 1987).

²⁷⁹ *People v. Linscott*, 566 N.E.2d 1355, 1362-63 (Ill. 1991) (internal citations omitted).

cerning the serology, found that the statements made by the prosecutor concerning the hair evidence alone did not warrant relief: “The prosecutor’s ‘match’ comments were improper, but we do not find, as did the appellate court, that they were so ‘egregious’ as to deny defendant a fair trial.”²⁸⁰

In other cases, courts did not provide any remedy at all. In the Larry Ollins case, the prosecutor—after stating that Ollins’s blood and hair shared characteristics with evidence found at the crime scene—referred to Ollins as a man “whose semen matches what’s found at the scene. Whose hair is found on the seat of the car.”²⁸¹ When the issue was raised on appeal, the court denied relief, emphasizing that “[a] prosecutor is permitted a wide range of expression in comments made during closing argument. Reversal will not be warranted even if the prosecutive comments could be viewed as improper, unless the defendant is substantially prejudiced.”²⁸²

In the Drew Whitley case, the criminalist, Dorothy Menges, the Forensic Serology Section Laboratory Manager at the Allegheny County Crime Laboratory, examined a number of very short shaved or cut hair fragments found on a stocking apparently worn by the perpetrator as a mask and found at the crime scene. Menges, though finding similarities despite admitting the hair fragments were unsuitable for comparison, was clear: “Because these hair fragments were so small, I could not make the statement that they were microscopically consistent.”²⁸³ During his closing argument, Assistant District Attorney Nicholas Radoycis extended Menges’s already invalid testimony, stating, “She said all the characteristics overlap, came from the same individual.”²⁸⁴ Radoycis

²⁸⁰ Id. at 1361.

²⁸¹ Trial Transcript at 62, *People v. Larry Ollins*, No. 02 L 000749 (Ill. Cir. Ct. June 20, 1988).

²⁸² *People v. Ollins*, 601 N.E.2d 922, 925 (Ill. App. Dist. 1992) (citation omitted). The Brian Piszczek case provides another example regarding serology evidence. The analyst properly explained that any male could have been the perpetrator where the stains exhibited the same A type as the victim and Piszczek was a non-secretor. Trial Transcript at 158, *State v. Brian Piszczek*, No. 257813 (Ohio Ct. Com. Pl. June 26, 1991). Yet the prosecutor argued in his closing argument, “What it is consistent with, ladies and gentlemen, that the person who did it would have been a non-secretor and who was a non-secretor? This man.” Id. at 260.

²⁸³ Transcript, Whitley, *supra* note 183, at 898. Menges never concluded that there was “no doubt” the hairs came from Whitley, as the prosecutor claimed.

²⁸⁴ Id. at 43.

then added, “Let’s see what the Crime Lab says. Dorothy Menges, do you remember the last thing she said on that witness stand Friday afternoon? I have no doubt those hairs came from Drew Whitley.”²⁸⁵ In fact, she had specifically rejected that conclusion. When the court asked her at the close of her testimony, “You can’t say it belongs to the defendant,” she had answered, “That is correct.”²⁸⁶ Despite stating the opposite of her actual testimony, Radoycis embellished further, telling the jury: “But it’s only when the scientists come in and say, hey, we have a standard, we know this hair to be of Drew Whitley and they compare it all microscopically. Exact. No doubt about it. (Pointing.) Him.”²⁸⁷ In response, the defense attorney brought a motion for mistrial:

The District Attorney stated to this jury that Dorothy Menges testified under oath that these hairs that she was comparing came from Drew Whitley. That is absolutely, positively not the evidence; and that is the most vital part of this whole case; and for him to say that constitutes prosecutorial misconduct. . . . She never said that they came from my client, Your Honor.²⁸⁸

Judge Walter R. Little equivocated, stating, “I do recall that she answered my question as she couldn’t say exactly who those hairs belonged to. . . . I don’t know if she did say it. I don’t recall.” When the prosecutor claimed he did hear such a statement and asserted that “[i]t’s the jury’s responsibility to remember things,” Judge Little provided a curative instruction that told the jury to resolve any discrepancy themselves.²⁸⁹

Each of these examples suggests the importance not just of accurate forensic testimony at trial, but the importance of the defense objecting and the court providing curative instructions should the

²⁸⁵ Id. at 45.

²⁸⁶ Id. at 950.

²⁸⁷ Id. at 50.

²⁸⁸ Id. at 51–52.

²⁸⁹ Id. at 53–55. The judge instructed:

I recall asking Mrs. Menges whether or not if she could say who the hairs in the stocking cap belonged to. It is my recollection she said that she could not. There has been some discrepancy as to whether or not Miss Menges could identify who the hairs in the stocking cap belonged to. Again, I want to caution you it will be your recollection which will prevail as to what Dr. Menges’ testimony is along with all the other testimony in this particular case.

Id. at 56–57.

science be presented in a misleading manner during closing arguments. Post-conviction courts are unlikely to provide a legal remedy for such conduct.

2. Failures of Defense Counsel

Defense 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. Defense counsel often also failed to address forensic science during their closing arguments. Defense experts testified in only 19 trials amongst the transcripts reviewed. For example, Gerald Davis presented an expert who contradicted Fred Zain's invalid testimony: "It is an absolute medical certainty that this seminal fluid did not come from Gerald Wayne Davis."²⁹⁰ That expert, Echols A. Hansbarger, Jr., noted that if he dismissed ABO exclusions as false results, "I'm afraid that I wouldn't be in my job very long."²⁹¹ He was paid \$100 for his work in the case.²⁹²

Stephen Linscott presented an expert who, on cross-examination, rejected any notion that probability of association can be used to evaluate hair evidence. The prosecutor nevertheless asserted that the defense expert endorsed such probability testimony, which the appellate court in reversing the conviction found to be a "calculated, rank misrepresentation."²⁹³ Curtis McCarty also presented an expert, John Wilson, the Chief Forensic Chemist of the Regional Crime Lab in Kansas City. Wilson only reviewed Joyce Gilchrist's report and did not conduct any independent analysis, but he corrected Gilchrist's faulty use of serology population statistics.²⁹⁴

Perhaps defense attorneys cannot be expected to understand scientific evidence and effectively cross-examine state experts, much less test the accuracy of the underlying data, without access to defense experts. Nevertheless, courts frequently deny the defense

²⁹⁰ Transcript, G. Davis, *supra* note 132, at 326.

²⁹¹ *Id.* at 331.

²⁹² *Id.* at 334.

²⁹³ *People v. Linscott*, 511 N.E.2d 1303, 1309 (Ill. App. Ct. 1987).

²⁹⁴ *McCarty v. State*, 765 P.2d 1215, 1218 (Okla. Crim. App. 1988) ("Mr. Wilson testified that Gilchrist's forensic report reflected that none of the pubic hairs found on the victim were consistent with appellant.").

funding for experts in criminal cases in which forensic evidence plays a central role.²⁹⁵ The presentation of forensic science testimony is typically one-sided in the majority of states that do not routinely fund the provision of forensic experts for indigent defendants. Moreover, in cases where defendants are able to present expert testimony, the experts are sometimes inexperienced or ineffective, and they may not have access to the underlying forensic evidence. Thus, it should come as no surprise that, despite the stakes, the defense does not often meaningfully challenge invalid forensic science testimony.

3. *Judicial Rulings on Forensic Science*

Courts policed the introduction of forensic testimony in these trials in a highly deferential manner, typically trusting the jury to assess the expert testimony.²⁹⁶ Defense attorneys did not often raise trial motions or objections regarding forensic testimony. In the Glen Woodall case, the defense moved to exclude Zain's "erroneous" serology chart because Zain had included a false statistic, dividing his figure in half and supposedly limiting his numbers just to men. The Court denied the motion: "That's in dispute. That's something the jury will have to determine."²⁹⁷ In Edward Honaker's case, the court denied a new trial motion, stating, "In the opinion of the court the evidence against you was overwhelming. You couldn't get around the scientific evidence that one of your hairs was found on her person."²⁹⁸

Similarly, fairly few of these exonerees challenged the forensic evidence during their appeals or post-conviction. Few among the first 200 people exonerated by post-conviction DNA testing whose

²⁹⁵ See 1 Giannelli & Imwinkelried, *supra* note 78, §§ 4.01–4.05, 13.07; Jack B. Weinstein, *Speech, Science, and the Challenge of Expert Testimony in the Courtroom*, 77 *Or. L. Rev.* 1005, 1008 (1998) ("Courts, as gatekeepers, must be aware of how difficult it can be for some parties—particularly indigent criminal defendants—to obtain an expert to testify.").

²⁹⁶ See Gross & Mnookin, *supra* note 18, at 169 ("Once a witness has been permitted to testify as an expert under Rule 702, judges usually leave the task of correcting and explaining their instructional statements to the opposing parties and the expert witnesses they call.").

²⁹⁷ Trial Transcript at 1168, *State v. Glen Dale Woodall*, No. 87-F-46 (W. Va. Cir. Ct. July 7, 1987).

²⁹⁸ Transcript, Honaker, *supra* note 64, at 29.

cases were examined in the first author's *Judging Innocence* study had asserted claims challenging the forensic evidence during their appeals or post-conviction, though 6 exonerees had such claims granted.²⁹⁹

Among those exonerated since that study was completed, Kennedy Brewer and Curtis McCarty brought claims related to the forensic evidence. In Brewer's case, the Mississippi Supreme Court twice rejected his claims challenging Michael West's erroneous bite mark testimony.³⁰⁰

In contrast, McCarty had his conviction reversed twice based on challenges to Joyce Gilchrist's testimony. First, McCarty had his 1986 conviction and death sentence reversed in 1988 due to Joyce Gilchrist's misconduct concerning the forensic analysis at his criminal trial. Regarding her agreement with the statement that "McCarty was physically present during the time violence was done to [the victim]," the court noted, "We find it inconceivable why Ms. Gilchrist would give such an improper opinion, which she admitted she was not qualified to give."³⁰¹ McCarty was convicted again in 1989, and Gilchrist's testimony at his second trial was found not to be grounds for reversal on appeal.³⁰² When his conviction was ultimately vacated again in 2005 based on Gilchrist's alteration and fabrication of laboratory reports, the court emphasized that "Ms. Gilchrist's actions alone warrant a new trial."³⁰³ In 2007, after serving twenty-two years in prison, McCarty was finally exonerated by post-conviction DNA testing.³⁰⁴

Courts denied relief to exonerees who asserted claims regarding some of the most flagrantly invalid forensic testimony reviewed in this study. Courts typically deferentially review any trial court evidentiary rulings as to the relevance of the proffered testimony, and

²⁹⁹ See Garrett, *supra* note 7, at 73–94.

³⁰⁰ See *Brewer v. State*, 819 So. 2d 1169, 1176 (Miss. 2002); *Brewer v. State*, 725 So. 2d 106, 134 (Miss. 1998).

³⁰¹ *McCarty v. State*, 765 P.2d 1215, 1218 (Okla. Crim. App. 1988).

³⁰² *McCarty v. State*, 904 P.2d 110, 125–29 (Okla. Crim. App. 1995) (denying relief on claims regarding Gilchrist's hair comparison testimony, but remanding for new capital sentencing hearing).

³⁰³ *McCarty v. State*, 114 P.3d 1089, 1094 (Okla. Crim. App. 2005) (affirming grant of new trial based on Gilchrist's *Brady v. Maryland* violations).

³⁰⁴ Cheryl Camp, *Convicted Murderer Is Freed in Wake of Tainted Evidence*, N.Y. Times, May 22, 2007, at A16; The Innocence Project, Curtis McCarty, <http://www.innocenceproject.org/Content/576.php> (last visited Nov. 10, 2008).

harmless error doctrines may further lead a court to excuse invalid admission of such testimony. Examples are easily found in reported cases outside this study set: for instance, though some courts have ruled that probability estimates by hair examiners are inadmissible,³⁰⁵ other courts have affirmed their use even in cases involving wildly fabricated numbers like a one in fifteen billion chance of a random hair match.³⁰⁶ One example from a case studied here is the testimony of Fred Zain in the Gerald Davis case, in which Zain testified that bacteria could account for serological results that excluded Davis. The Supreme Court of Appeals of West Virginia noted that Zain explained the presence of characteristics foreign to Davis as “the result of a false reading due to bacterial contamination.”³⁰⁷ Given a forgiving sufficiency standard in which the evidence is viewed in the light most favorable to the prosecution, the court found that no “injustice ha[d] been done,” and that “the scientific evidence d[id] not conclusively demonstrate the appellant’s innocence.”³⁰⁸

Similarly, in the Larry Ollins case, the Appellate Court of Illinois denied relief despite the testimony of Pamela Fish, an expert from the Chicago Police Department Crime Laboratory, who falsely asserted that “the defendant’s blood sample matched semen found in the victim.” The court observed that “the test results were

³⁰⁵ See, e.g., *United States v. Massey*, 594 F.2d 676, 680 (8th Cir. 1979) (“There is no foundation to show the factual circumstances surrounding each of [the expert’s] examinations and certainly there is no statistical probability which could be drawn from his experience to show that there was only ‘one chance in a 1,000’ that hair comparisons could be in error.”); *State v. Carlson*, 267 N.W.2d 170, 175–76 (Minn. 1978) (finding hair comparison testimony to be “improperly received” where it provided a “suggestion of mathematical precision”; Barry Gaudette had testified that, based on his study, he found a “1-in-4,500 chance that the head hairs did not belong to the accused”); *Butler v. State*, 108 S.W.3d 18, 21–27 (Mo. Ct. App. 2003); *State v. Scarlett*, 426 A.2d 25, 28–29 (N.H. 1981); *State v. Faircloth*, 394 S.E.2d 198, 202–03 (N.C. Ct. App. 1990).

³⁰⁶ See, e.g., *Lopez v. State*, 643 S.W.2d 431, 433 (Tex. App. 1982) (“This expert testified that the chances of the resemblance he found between the hair samples occurring at random was one in 1.5×10^{10} (1 in 15,000,000,000).”); *State v. Bridges*, 421 S.E.2d 806, 808 (N.C. Ct. App. 1992) (finding harmless error where expert testified that the “likelihood of two Caucasian individuals having indistinguishable head hair . . . is very low. A conservative estimate for that probability would be . . . approximately one in a thousand.”).

³⁰⁷ *State v. Davis*, 376 S.E.2d 563, 567 (W. Va. 1988).

³⁰⁸ *Id.* at 568.

corroborated by an eyewitness account of the crime. Additionally, the State points out that the jury was instructed as to the specific limitations of the test results in both the opening and closing arguments of the prosecution.” The court concluded, “Because the test results were corroborated by substantial independent evidence, we find that Fish’s testimony was properly admitted into evidence by the trial court.”³⁰⁹

The Montana Supreme Court in the Chester Bauer case found no reversible error in analyst Julie Long’s testimony. Long had ignored the problem of masking and quantification and falsely divided the statistic offered in half, claiming to rule out females. Presented with a claim regarding this invalid testimony, the court denied relief, stating that “the fact remains that Bauer is a secretor, and that should be relevant.”³¹⁰

B. A National Framework for Reform of Forensic Science

In 1989, just as DNA technology arrived, Eric Lander commented, “At present, forensic science is virtually unregulated—with the paradoxical result that clinical laboratories must meet higher standards to be allowed to diagnose strep throat than forensic labs must meet to put a defendant on death row.”³¹¹ Two decades later, that state of affairs has changed little, making the invalid testimony prevalent in these wrongful conviction cases unsurprising. No legislation or oversight mechanisms regulate the quality of forensic science reports or testimony.

DNA exonerations have provided some impetus for state and local forensic science reform. Several states have enacted legisla-

³⁰⁹ *People v. Ollins*, 601 N.E.2d 922, 924–25 (Ill. App. Ct. 1992); see also *People v. Saunders*, 603 N.E.2d 32, 36 (Ill. App. Ct. 1992) (“Fish’s testimony is corroborative of the defendant’s own admission that he was present . . . even though the semen samples taken from the victim excluded the defendant.”).

³¹⁰ *State v. Bauer*, 683 P.2d 946, 951 (Mont. 1984). Compounding the problem, the court supported its ruling with reference to additional invalid science testimony:

[T]here is other independent evidence of Bauer’s guilt. . . . Arnold Melnikoff, Bureau Chief of the State Crime Laboratory, testified that pubic hair and head hair found at the crime scene were similar to Bauer’s pubic and head hair. Melnikoff estimated that the chances of another person having the same type of pubic and head hair were one in ten thousand.

Id.

³¹¹ Eric S. Lander, DNA Fingerprinting on Trial, 339 *Nature* 501, 505 (1989).

tion creating independent bodies to review their crime laboratories in response to misconduct.³¹² However, very few exonerations have resulted in scrutiny and audits of the invalid science that supported the wrongful conviction. Regarding DNA laboratories, Congress, in 1993, with the establishment of a national DNA databank, created a DNA Advisory Board to provide quality assurance standards.³¹³ Many state and local crime labs voluntarily participate in ASCLD/LAB.³¹⁴ But voluntary programs run by the crime laboratory directors, although a positive step, fail to address the needs identified in this study. None deal with the claims made by forensic analysts in their reports or in their trial testimony.

For disciplines other than DNA analysis, the federal government in 2004 required states to create an entity and process for external independent audits to respond to allegations of misconduct or serious negligence in laboratories receiving federal funds.³¹⁵ That legislation, however, has not been enforced, according to the Inspector General, and many states are not in compliance.³¹⁶ Moreover, the audits are limited to misconduct in the government laboratories and thus fail to cover the serious misconduct of unaffiliated forensic analysts hired by prosecutors, including forensic odontologists and the employees of police departments that conduct analyses of ballistics and fingerprint evidence.

Despite these faint stirrings of reform, crime laboratories and forensic analysts remain remarkably free from oversight and still lack basic scientific standards to govern their conclusions. No federal

³¹² See, e.g., Minn. Stat. § 299C.156 (2007) (establishing Forensic Laboratory Advisory Board); N.Y. Exec. Law §§ 995-a to -b (McKinney 2003) (establishing forensic science commission and requiring accreditation); Okla. Stat. Ann. tit. 74, § 150.37 (West 2007) (requiring accreditation); Tex. Code Crim. Proc. Ann. art. 38.35(d) (Vernon 2005) (requiring accreditation by the Texas Department of Public Safety); Va. Code Ann. § 9.1-1101 (2006) (creating Department of Forensic Science and oversight committee).

³¹³ 42 U.S.C. § 14132(a), (c) (2000).

³¹⁴ See ASCLD/LAB Home Page, <http://www.asclcd-lab.org/> (last visited Nov. 10, 2008).

³¹⁵ See 42 U.S.C. § 3797(k)(4) (Supp. IV 2007) (requiring that laboratories receiving federal grants create mechanisms for external independent investigations).

³¹⁶ See Oversight of the Department of Justice's Forensic Grant Programs: Hearing Before the S. Comm. on the Judiciary, 110th Cong. (2008) (statement of Glenn A. Fine, Inspector General, U.S. Department of Justice); Oversight of the Justice For All Act: Hearing Before the S. Comm. on the Judiciary, 110th Cong. (2008) (statement of Peter Neufeld on Behalf of The Innocence Project).

legislation regulates the quality of non-DNA forensic disciplines or the content of reports or testimony, which is significant because the overwhelming majority of crime lab work involves techniques other than DNA analysis.

Nationally enforceable standards should govern the interpretation of forensic data within scientifically acceptable parameters. The authors also appreciate the need for a federal infusion of capital for both basic and applied research and to ensure that only validated and reliable methods and assays are used to analyze evidence of crime. But since these are needs that are not derived solely from this study set, they will not be addressed here.³¹⁷

National scientific standards should address the use of forensic science: both the methodology and, importantly, the way that conclusions are drawn from evidence. Even in areas such as ABO blood typing, in which there is consensus on the reliability of the testing methods, invalid testimony can result from a failure to ensure adherence to scientific standards when drawing potentially unsound conclusions from sound testing methods. In contrast, invalid science testimony was unsurprising in disciplines where there was simply no consensus on the boundaries of permissible trial testimony. Even in disciplines that provided non-binding guidelines, no criteria were provided for reaching conclusions. Indeed, bite mark and shoe print guidelines explicitly permit conclusions unsupported by science. Furthermore, the forensic disciplines have created no means to enforce any scientific standards.³¹⁸ Forensic laboratories have also not created effective mechanisms for review of analysts' work and have typically not responded even after invalid forensic testimony and analysis was uncovered. Outside intervention is urgently needed.

The NAS Committee report provided the long overdue recommendation that we meaningfully regulate the presentation of forensic science. In particular, the NAS report recommended that

³¹⁷ See, e.g., Giannelli, *supra* note 32, at 72–76, 87–89 (discussing proficiency testing, accreditation of crime laboratories, and other avenues of reform); Henry C. Lee, *Forensic Science and the Law*, 25 Conn. L. Rev. 1117, 1124 (1993) (“Perhaps the most important issue in forensic science is the establishment of professional standards.”).

³¹⁸ See Gross, *supra* note 62, at 1178 (“Unfortunately, what an expert says in court is generally invisible and inaudible in her own professional world. If expert witnesses were accountable to their colleagues, even informally, they might fear the consequences of irresponsible testimony far more than they do.”).

Congress establish an independent federal agency, a “National Institute of Forensic Science,” which would promulgate “standard terminology” for report writing and testimony.³¹⁹ The history of the development of standards for DNA analysis provides a model. The National Institutes of Health funded basic early research, as did universities and other institutions. The National Institute of Standards and Technology promulgated quality assurance standards, including match criteria. A National Institute of Forensic Science, led by independent scientists—not those employed by crime laboratories—could take on the important task of developing scientific criteria for interpreting data in many of the non-DNA disciplines, particularly those that attempt to identify a person or object in connection with evidence recovered from a crime scene. The forensic disciplines would benefit from scientific criteria to promote consistency and best practices. Invalid written reports and expert testimony damage the credibility of the entire forensic science community. National standards would reduce the number of wrongful convictions and enhance the likelihood that forensic science could help identify the real perpetrator.

The NAS Committee’s report and the responses of those in the scientific community will contribute to a national discussion regarding the future of the forensic sciences. That discussion will hopefully lead to the type of legislation and oversight proposed. Forensic sciences urgently require a far more rigorous scientific foundation. Specifically, there should be a sound foundation for the process by which analysts reach the conclusions in their reports and trial testimony. Should reformers focus only on methodology and not also on the actual practice of forensic science in the courtroom, invalid testimony and miscarriages of justice will continue to tax our criminal justice system and society.

CONCLUSION

This exploration of the types of invalid forensic science testimony that have contributed to convictions of the innocent provides one window into failures of our criminal justice system to adequately review the use of forensic science. That system still has not responded with a full investigation into most of these known mis-

³¹⁹ See Strengthening Forensic Science, *supra* note 12, at S-14, 6-3-6-5.

2009]

Invalid Forensic Science Testimony

97

carriages of justice, much less looked into other potentially affected cases or routinely conducted investigations to ensure accurate dispositions. These cases suggest that the adversary system cannot be depended upon as an adequate safeguard. The defense bar lacked the expertise and resources to detect and address invalid forensic science effectively in most of these cases, and judges did not remedy most errors brought to their attention.

Finally, the invalid science testimony described here ranges from cases in the 1980s involving conventional forensic disciplines employing visual comparison, to serology analysis employing clear population statistics, to the use of modern DNA technology in the 1990s. Though the technology has changed over time, the sources of human error, misinterpretation, and misconduct have not. This body of cases in which innocent people were convicted based on invalid forensic testimony demonstrates the importance of efforts to study the validity of forensic science as presented in reports and testimony at trial and encourages us to rethink how our criminal system handles forensic science in the laboratory and in the courtroom.

The evidence from these wrongful conviction cases supports efforts to create and enforce national standards for interpreting data in the various forensic disciplines. The scientific community can take the lead in reform efforts. Detailed scientific standards are needed to establish permissible parameters both for report writing—particularly important since so many cases are resolved by plea bargains—and for trial testimony. As the criminal trials of these innocent people demonstrate, if reforms are not implemented to regulate the practice of forensic science during criminal investigations long before the adversary process begins, the opportunity to avert miscarriages of justice may be lost.

2010 WL 2539342 (N.D.Ill.) (Expert Report and Affidavit)
United States District Court, N.D. Illinois,
Eastern Division.

UNITED STATES OF AMERICA,
v.
Clacy Watson HERRERA.

No. 01 CR 1098.
March 8, 2010.

Expert Disclosure Report of Lyn Haber, Ph.D. and Ralph Norman Haber, PH.D.

Case Type: Criminal >> Drugs/Narcotics

Jurisdiction: N.D.Ill.

Name of Expert: [Lyn Haber, Ph.D.](#)

Area of Expertise: Forensics >> Finger Print/Latent Examiner

Representing: Defendant

Name of Expert: [Ralph Norman Haber, Ph.D.](#)

Area of Expertise: Forensics >> Finger Print/Latent Examiner

Representing: Defendant

I. BACKGROUND

A. Human Factors Consultants

Lyn Haber, Ph.D. and Ralph Haber, Ph.D. are joint partners of Human Factors Consultants, a firm that has provided consultation, research and expert testimony to the legal profession on identification matters since 1980. We are both research scientists specializing in the evaluation of research and practice of forensic identifications, and in the underlying cognitive processes of perception, memory, decision-making, judgment, language, comprehension, training and reporting.

B. Qualifications of Lyn Haber, Ph.D.

Lyn Haber has four degrees, including a BA from Brandeis University, a MA and Ph.D. from the University of California at Berkeley, and an MA from the University of Illinois at Chicago. She has taught graduate and undergraduate science courses at Temple University, the University of Rochester, Stanford University and the University of Illinois, rising from the rank of Instructor to Full Professor. She is presently a partner in Human Factors Consultants. She has published 150 scientific articles and books, and presented at 150 scientific conferences. She has been retained in over 200 criminal and civil litigation cases, and has testified in both federal and state superior courts. Her resume is attached as Appendix A.

C. Qualifications of Ralph Haber, Ph.D.

Ralph Haber has a BA from the University of Michigan, a MA from Wesleyan University, a Ph.D. from Stanford University, and postgraduate training from the Medical Research Council at Cambridge, England. He has taught graduate and undergraduate courses at San Francisco State College, Yale University, the University of Rochester, Stanford University, and the University of Illinois, rising from the rank of assistant professor to full professor. He was chairman of the Department of Psychology at the University of Rochester. He is presently a partner in Human Factors Consultants. He has published 250 scientific articles and books, and presented his work at 150 scientific conventions and conferences. He has been retained in over 200 criminal and civil litigation cases, and has testified in both federal and state superior courts. His resume is attached as Appendix B.

D. Research Publications and Conference Presentations on Fingerprint Science

From 2001 to the present, we have jointly published one book and six articles on fingerprint science. Our book “Challenges to Fingerprints” (2009) details the scientific lacks in forensic fingerprint comparison and suggests specific research experiments to remedy each unknown. The articles include a chapter for the California Bar Association and a court brief challenging the admission of fingerprints; the remainder are research papers. We have given nineteen presentations, of which nine were invited: two to university audiences, two to a conference of public defenders, three to examiners working in police and sheriff crime laboratories, one to a conference of private investigators, and one to the California Commission on the Fair Administration of Justice. The remaining ten presentations included five at scientific conventions, two as part of continuing legal education training, and three to Innocence Project trainees and lawyers. A complete listing of these scientific publications and presentations is attached as Appendix C.

E. Testimony at Trial on Fingerprint Cases

In this same time period, we have been retained on 27 cases (11 in federal courts). Of those, we provided consultation only on fingerprint science (8 cases); we have submitted briefs, affidavits or reports in 8 cases; and we have testified (including the present case) in 11 cases, 5 in state courts, including 4 Frye hearings; and 6 in federal courts, including 2 Daubert hearings. When we have submitted affidavits or briefs, or have testified, the court has ruled in every case that we were qualified to provide testimony and/or documentation on fingerprint science. A complete listing of these cases is attached as Appendix D.

F. Training in Fingerprint Comparisons

We have both taken three training courses in latent fingerprint comparison since 2003, offered by highly trained, certified IAI fingerprint examiners. Two were under the auspices of the IAI post-graduate training programs. These comprised 104 classroom hours for each of us. The courses focused on the method for performing latent print comparisons, the scientific issues underlying fingerprint identifications, and training and proficiency testing of latent fingerprint examiners. This classroom training on comparisons exceeds the amount received by an average full-time fingerprint examiner working in a crime laboratory.

G. Familiarity with Research Publications on Fingerprint Comparisons

The vast majority of articles and books published since the end of the 19th century on fingerprint comparison procedures and evidence have been written by fingerprint comparison technicians trained to offer opinions on whether two fingerprints match or not. With very few exceptions, these publications neither present scientific evidence of the accuracy of comparison processes nor discuss questions of accuracy. We are familiar with most of these books and articles.

In contrast, we have read and carefully reviewed the entire (though not large) scientific literature on experimental research on fingerprint comparisons. We wrote the first article ever published on evidence of the error rate of fingerprint comparisons. We also wrote the first article ever published on the scientific evidence required to respond to the Daubert criteria for court acceptance (attached as Appendix E).

II. INTRODUCTION

A. Patterns on Fingers and Fingerprints.

The surfaces of human fingers (and palms, toes and soles) are covered with friction ridge skin made up of intricate patterns of ridges and grooves to increase friction with surfaces. For centuries, it has been assumed that such patterns on friction ridge skin are unique to each finger of each person, living, dead, or unborn. However, fingers are not used during the comparison process, since perpetrators of a crime never leave their fingers at the crime scene. Rather, when a person touches a surface, and either his fingers and/or the surface has sufficient moisture, the pattern of ridges and grooves on the finger leaves an impression image on the surface. This image is called a fingerprint, and fingerprints are the stimuli for the forensic science of fingerprint comparison and identification.

The forensic hypothesis for fingerprint comparison states that if two fingerprints have the identical pattern, the two fingerprints were made by the same finger (person). On this basis, a forensic science of fingerprint comparison was founded in the 19th century, in which a person could be identified by his fingerprint, and it was assumed that there was no chance that a fingerprint could match more than a single person.

The images of the patterns of friction ridge skin are not identical to the patterns on fingers themselves. The friction ridge skin on a finger is three dimensional, whereas images of fingers are two dimensional and therefore distorted (as a flat map distorts the shapes of the land masses shown on a globe). Further, because the friction ridge skin on the surface of the finger is flexible, the pressure of the touch (heavy or light) changes the shape of the finger, and therefore the widths of ridges and grooves and their relative distances from each other. If the pressure is not straight down, but slides along the surface, or rotates, the pattern of ridges and grooves is altered. The matrix (gooey or thin) on the surface of the finger that transfers the image also changes the pattern. The shape of the surface on which the print is deposited, the presence of noise on that surface (such as dirt or other fingerprints) also alters the pattern. The process used to lift the fingerprint and to enhance its visibility also alters the pattern. Each of these factors produces distortion in the image that is not part of the pattern from the friction ridge skin on the finger. For these reasons, *every fingerprint made by a particular finger differs from every other fingerprint made from the same finger, and differs from the pattern on the finger itself.* The target stimulus for a fingerprint comparison is not the pattern on the friction ridge skin of a finger, but is a distorted image of that pattern.

Distortion can be minimized whenever great care is used in making a fingerprint, as is exercised by a trained technician in a police station. These are referred to as exemplar fingerprints.

When a finger touches a surface inadvertently, this is called a latent fingerprint. It is an uncontrolled touch, only part of the finger's friction ridge skin comes into contact with the surface, and there is no control over the pressure or other causes of distortion. A latent fingerprint is typically much smaller than an exemplar fingerprint (on average only 1/5 the size), and contains far less information content, and frequently is heavily distorted. When a perpetrator touches a surface at the crime scene, he leaves a latent fingerprint.

B. The Task of a Fingerprint Examiner.

An examiner is given an unknown latent fingerprint and a carefully made inked or scanned exemplar fingerprint of a suspect or known innocent. The examiner is asked to apply a method to compare the unknown latent to the known exemplar fingerprint, and to offer an expert opinion as to whether the two fingerprints were made by the same donor finger or by two different donor fingers. If the opinion is one donor, then the donor of the unknown latent can be identified to the donor of the known exemplar fingerprint. If the opinion is two donors, then the donor of the unknown latent print remains unknown, since s/he is not the donor of the known exemplar fingerprint.

C. Fingerprint Examiners Use the ACE Method of Fingerprint Comparison.

The fingerprint profession states in its publications that fingerprint examiners use the Analysis-Comparison-Evaluation method (ACE) (e.g., SWGFAST, 2002a). When a second examiner confirms the conclusion of the first examiner, this is called Verification, and the method is referred to as ACE-V. These are the only methods we consider in this affidavit.

What follows is a very general framework of the ACE method we have extracted from fingerprint examiner publications.

The Analysis Stage concerns the latent fingerprint. The examiner first determines whether it contains enough quality and quantity of information to compare to exemplar fingerprints without serious risk of error. If the latent passes this value test, the examiner analyzes the physical evidence contained in the latent print, such as the surface on which it was left (e.g., smooth, irregular, porous, dirty), the distortions that altered it from the finger (e.g., heavy downward or sideways pressure), the orientation of the print, the most likely finger used, and any special problem areas in the latent that cannot be used for comparison. The examiner then notes the important ridges and/or features that can be used to start comparison, locating each one spatially in relation to other features in the latent.

The Comparison Stage requires the examiner to look from the ridges and features he labeled in the latent print to determine whether each one can be found in the exemplar in comparable locations. If he finds most or all of them, then he selects further features in the latent and then looks again at the exemplar to determine whether these further features are present in the corresponding locations. If the comparisons between the latent and the exemplar fail to show correspondence of feature type and location, the examiner excludes the exemplar print (and the suspect whose finger made that print). If the examiner concludes that all of the differences in correspondence are due only to distortion, and not to discrepancies that arose because the exemplar finger did not make the latent print, the examiner proceeds to evaluation.

The Evaluation Stage requires the examiner to total up the amount of similarity between the two fingerprints and determine whether it is sufficient to opine that the two prints were made by the same finger. If he concludes that there is a sufficient amount of similarity, he concludes an identification. If the similarity was not sufficient, he concludes insufficient or inconclusive.

The Verification Stage, when performed, may be carried out in two ways. In the non-blind, typical procedure, another examiner ratifies the work and conclusion of the first examiner (the ACE-V version). When the examination is repeated blindly, the second examiner repeats the entire ACE in ignorance of the fact it has been carried out previously. This procedure is more properly called ACE-ACE.

D. Assessment of the Accuracy (Validity) of ACE-Based Conclusions.

The accuracy of the expert opinion proffered by the latent fingerprint examiner depends upon the proficiency of the examiner in applying the comparison method, and on the power of the method to produce correct conclusions when properly applied.

Scores obtained on a validated proficiency test containing test materials comparable in type and difficulty to normal casework assess the proficiency of fingerprint examiners.

The power of the method to produce correct conclusions is assessed by demonstration of the method's validity. In science, validity refers to the accuracy of conclusions based on the method, compared with ground truth (knowledge of the true state of the facts: for fingerprint comparisons, ground truth refers to whether the two prints were made by one donor or two donors). The validity of the ACE method to compare fingerprints is the accuracy of conclusions reached by fingerprint examiners who are fully trained and experienced in using the ACE method, comparing latent-exemplar pairs typical of casework but for which ground truth is known, working under optimal conditions. Any errors made under these circumstances would result from failures of the ACE method itself to produce the correct decision as to one or two donors. The validity of a method such as ACE is

usually reported as the converse of its accuracy-as an error rate. If the experiment just described showed that the method results in correct conclusions 95% of the time, the method's error rate would be 5%.

III. OPINIONS

A. OPINIONS REGARDING THE UNIQUENESS ASSUMPTIONS OF PATTERNS ON FINGERS AND FINGERPRINTS.

The fingerprint profession correctly asserts that if fingerprints are to be used to identify people, friction ridge patterns on the skin must be permanent and must be unique to each finger (SWGFAST, 2002a, Ashbaugh, 1999). We do not contest here either part of this assumption when it is applied exclusively to the friction ridge skin on fingers.

Based on the uniqueness assumption and professional requirements, fingerprint examiners testify to an “absolute” identification: “The latent print found at the crime scene was made by the same person who made the exemplar prints used for comparison, to the exclusion of every other person” (Ashbaugh, 2005, SWGFAST, 2002a).

A fingerprint examiner compares fingerprints, not fingers, so a uniqueness assumption is needed for exemplar fingerprints, and for latent fingerprints. If an exemplar fingerprint is unique, it implies two things. First, (1) its underlying pattern is unique to the finger that made it. An equivalent statement is that its pattern can only be matched to the finger that made it. Because the friction ridge skin on the finger is flexible, every exemplar print made by that finger will differ - they cannot simply be overlaid. Despite these differences, the basic pattern in every exemplar from a single finger must be matchable back to the true donor finger and to no other. This is an exemplar-to-finger uniqueness assumption.

Second, (2) the patterns in this exemplar, and all other exemplars made by this finger, are never confusable with the patterns in an exemplar from some one else. This is the exemplar-to-exemplar uniqueness assumption. These two assumptions permit the conclusion that if the patterns in an *unknown exemplar* match those in a *known exemplar* from a suspect, that suspect's finger made the unknown exemplar.

If a latent fingerprint is unique, it means that (3) its underlying pattern is unique to the finger that made it. Said another way, the patterns in the latent can only be matched to the finger that made it. Because the friction ridge skin on the finger is flexible, every latent print made by that finger will differ - they cannot simply be overlaid. Despite these differences, the basic pattern in every latent (of value for comparison) from a single finger must be matchable back to the true donor finger and to no other. This is a latent-to-finger uniqueness assumption. Further (4) the patterns in this latent are never confusable with the patterns in an exemplar from somebody else. This is a latent-to-exemplar uniqueness assumption. Assumptions (2) and (4) permit the conclusion that if the patterns in an *unknown latent print* match those in a *known exemplar* taken from a suspect, that suspect made the unknown latent fingerprint.

Opinion 1. The Assumptions Underlying Forensic Fingerprint Comparison are Unsubstantiated.

It is our opinion that the uniqueness assumptions (2) and (4) have not been addressed by the fingerprint profession, have not been demonstrated empirically to be true, and are contradicted by empirical evidence that indicates that, to an unknown but statistically significant extent, exemplars from two different fingers are confusable as having been deposited by the same finger; and, to an unknown but statistically significant extent, a latent and exemplar deposited by two different fingers are confusable as having been deposited by the same finger. We have five bases for this opinion.

First Basis. The Prominent Fingerprint Texts Describe the Permanence and Uniqueness Assumptions for Fingers Only.

The most referenced and respected fingerprint texts assert that friction ridge patterns are unique and permanent, but do not refer to the uniqueness assumptions needed for exemplars and latents (Ashbaugh, 1999; Champod, et al., 2004; Conger, 1993; Moenssen, 1995; Olsen, 1978; and the FBI classic text, 1988).

Second Basis. No research Studies have Addressed the Exemplar-to-Exemplar or Latent-to-Exemplar Uniqueness Assumptions using Fingerprint Examiners.

No experimental results, using human examiners, have been published to support the basic assumptions that different takes of exemplar and latent fingerprints can be correctly matched to other images from the same finger, and only to images from the same finger. Haber & Haber (2009, chapter 2), and Cole et al. (2008) report attempts to find such studies in the published (and unpublished) research literature, and are confident none exist. The National Academy of Sciences (NAS) report (2009, section 5, p. 13) explicitly comments on this failure: “Uniqueness does not guarantee that prints from two different people are always sufficiently different that they cannot be confused, or that two impressions made by the same finger will also be sufficiently similar to be discerned as coming from the same source. The impression left by a given finger will differ every time, because of inevitable variations in pressure, which change the degree of contact between each part of the ridge structure and the impression medium. None of these variables - of features across a population of fingers or of repeated impressions left by the same finger - has been characterized, quantified, or compared.” The NAS report further states that without evidence of the uniqueness assumption applied to exemplar and latent fingerprints, the assumption is of no value to support individualization to the exclusion of all others. A fingerprint conclusion of identification can support only a probability statement that there is some chance the donor of the exemplar is also the donor of the latent fingerprint.

Third Basis. The Exemplar-to-Exemplar Uniqueness Assumption Failed in a Recent Study Using an Automated Search System.

Cole, et al., (2008) used an automated search system to score pairs of exemplars for their similarity and found that the exemplar-to-exemplar uniqueness assumption failed. If the exemplar to exemplar assumption holds true, the results in such a study would show that two different exemplars made from the same finger always produce high similarity scores and two exemplars made from two different fingers always produce low similarity scores. Cole et al. used a U.S. National Institute for Standards and Technology (NIST) database consisting of two sets of different ten-print exemplars (twenty impressions from each person) taken from 2,700 individuals (54,000 prints). The authors selected a subset of 6,750 mated prints and compared them using an AFIS system to the remaining 53,999 prints in the database. Each search requested the ten most similar candidates to the submitted latent. The authors found that the true match had the highest similarity score (appeared as the top candidate) in 75% of the comparison outputs. The true match ranked among the other candidates in 4% of the outputs. In the remaining 21% of the comparison outputs, the true match was not included in the top ten candidates: the true match has less similarity to the submitted latent than at least ten other non-matching exemplars.

These results show that some false matches received higher scores than the true matching exemplar. For twenty one percent of the comparisons, exemplars from ten other fingers received higher similarity scores than the true matching exemplar. In twenty five percent of the comparisons, these exemplars were more similar to exemplars from a different individual than to a different exemplar made from the same finger. To the extent that these data would be found if human examiners performed the comparisons, the exemplar-to-exemplar uniqueness assumption fails for 25% of exemplars.

Fourth Basis: The Latent-to-Exemplar Uniqueness Assumption Failed in a Recent Study Using an Automated Search System.

Cole et al (2008) used their AFIS to determine whether a latent and an exemplar from the same finger would always have a higher similarity score than a latent and exemplar from different fingers. The NIST database in their second study contained 258 latent-exemplar pairs, which were different takes of the same finger. The authors found that the true exemplar match for

the latent was ranked as the top candidate in 70% of the comparison outputs, and was ranked among the other nine candidates in 12% of the comparisons. The candidate list failed to contain the correct exemplar in about 18% of the comparison outputs. The authors reported that all the false matches received a higher score than the lowest scoring true match. Thirty-four percent of the true matches scored lower than the highest scoring false match. These data suggest that the uniqueness hypothesis fails for latent to exemplar comparisons: the pattern on a given latent fingerprint is confusable with exemplars from different donor fingers in 30% of the comparisons.

Fifth Basis: The Results of Meagher et al (1998) Fail to Assess Either Uniqueness Assumption.

The Cole et al. study replaces an oft-quoted, improperly conducted, and unpublished demonstration produced as part of a Daubert hearing on the admissibility of fingerprint evidence (Meagher, Budowle and Ziesig, 1998). Meagher et al. used an AFIS system to compare 50,000 exemplars with themselves, not a second take of the same finger. This was an *identity* comparison, and is irrelevant to the claim that two different exemplars from the same finger can always be matched to each other, and not to any other exemplar.

To test the latent uniqueness assumption, Meagher et al (1998) manufactured “latents” by masking off all but the center 21.7% of the 50,000 exemplars (the FBI’s estimate of the average size of a latent), and submitted them to AFIS for comparison with the original 50,000 exemplars from which they had been masked. Again, these results cannot be offered as evidence that latent fingerprints are unique, because no latent prints were tested. Latent fingerprints are not masked-off centers of identical exemplar fingerprints. A latent print of a finger is a different impression of the finger, not an identical one; latent fingerprints rarely contain the information-rich center; and usually their quality is far poorer than that of the exemplar.

Until other experiments are published, especially ones using fingerprint examiners and not computers, the evidence suggests that both exemplar and latent fingerprints are frequently confusable with exemplars from different fingers. Using the data from Cole et al (2008) as a basis, the latent-to-exemplar uniqueness assumption failed for 30% of the comparisons. Even if the ACE method were infallible and examiners applied it perfectly, its error rate could never be less than 30%, because the error is built into the stimuli being compared, not into the method.

Opinion 2. Absolute Conclusions are both Unscientific and Wrong.

Even if the ACE method were infallible and fingerprint examiners applied it perfectly, its error rate could never be less than 30% (based on the data from Cole, et al., 2008), because the confusability is built into the stimuli being compared, regardless of the method. This confusability produces an irreducible error rate in fingerprint comparisons, so that absolute conclusions are not only unscientific but demonstrably false. “Therefore, in order to pass scrutiny under Daubert, fingerprint identification experts should exhibit a greater degree of epistemological humility. Claims of ‘absolute’ and ‘positive’ identification should be replaced by more modest claims about the meaning and significance of a ‘match’ ”(Mnookin, 2008, quoted in the NAS, 2009, 5-12).

The president of the IAI, Robert J. Garrett, in response to the NAS, suggested that IAI “members not assert 100% infallibility (zero error rate) when addressing the reliability of fingerprint comparisons,” and adds, “...the IAI does not, at this time, endorse the use of probabilistic models when stating conclusions of identification...” (IDentification News, 2009, vol. 39, No. 2). The IAI has not yet suggested the alternative testimony examiners should give.

B. OPINIONS REGARDING THE ACE METHOD'S FAILURE TO MEET DAUBERT CRITERIA.

The U.S. Supreme Court’s Daubert ruling contains several criteria to be used to evaluate admissibility of evidence based on a scientific method. With respect to the ACE method used by fingerprint examiners, we consider five criteria: (1) the acceptance by the relevant scientific and professional communities of the ACE method used to produce the evidence; (2) a substantial body of published research on the ACE method in peer reviewed scientific journals; (3) evidence that the validity of the ACE

method is testable; (4) evidence from scientific tests of the reliability and validity of the ACE method; and (5) evidence of a known error rate of the conclusions produced by the ACE method. Our opinion, as research scientists, is that the ACE method fails all five of these criteria.

Opinion 1. Neither a Community of Research Scientists nor a Community of Fingerprint Examiners Accepts a Scientific Basis for the Validity of the ACE Method.

Two communities are relevant to acceptance of the ACE method. The first is comprised of research scientists who are trained to do research, have conducted and evaluated research, and have evaluated research on the acceptability of the ACE method as encompassing good science. The other is comprised of forensic practitioners who are trained and employed to carry out fingerprint comparisons using the ACE method. Neither community accepts the scientific bases necessary for validity of the ACE method.

First Basis: There is a Near-total Absence of Support of the Validity of Fingerprint Methodology by Research Scientists.

When research scientists specializing in scientific methodology have testified or contributed to affidavits or amicus briefs concerning the admissibility of fingerprint comparison evidence, they have uniformly testified that there is no evidence for the validity of the ACE method. In the 50+ hearings that have been held on the scientific acceptability of fingerprint evidence, counting both federal and state courts (Onin Website on Legal Challenges, 2009), not a single research scientist who specializes in validity research on comparison methods has testified for the government or state in support of the research showing the validity of the ACE method, or in the support of a known error rate for the ACE method. The only trained research scientists who have testified for the Government or State have been embryologists, who limited their testimony to the fetal development of friction ridge skin (e.g., Babler, 1991).

Second Basis: The National Academy of Sciences, One of the Most Valued Body of Scientists in the World, Criticized the Absence of Scientific Evidence for the Validity of ACE.

A recent report from the National Academy of Sciences (2009) on forensic science provided a thorough review of the scientific support for fingerprint comparison. That report commented on a number of inadequacies in evidence in support of ACE, including the absence of evidence of its validity: “ACE provides a broadly stated framework for conducting friction ridge analyses. However, this framework is not specific enough to qualify as a validated method for this type of analysis” (NAS, 2009, 5-12).

Third Basis: The Professional Community of Fingerprint Examiners Has Neither Specified Nor Accepted a Specific Version of the ACE Method.

The most powerful demonstrations of acceptance of ACE by the fingerprint profession would include: (a) its members create an official published description of the method; (b) the published method is taught in official training programs; (c) assessment of performance by proficiency and certification testing is anchored explicitly to the official description of the method; and (d) examiners demonstrate that they follow the method correctly by providing contemporaneous bench notes of each examination. None of these demonstrations obtain. We list the further bases for this opinion.

Fourth Basis: The Steps of ACE and their Sequence have not been Fully Specified by the Fingerprint Profession.

The lack of a complete description of the ACE method has received considerable attention. As examples, (a) Interpol created a working group in 2000 to define a method for fingerprint identifications and issued a report urging its member countries to

adopt a common method (IEEGFI II, 2005); (b) a recent FBI report included a demand for a complete description of the method (Smrz, et al., 2006); (c) SWGFAST (2002) has a working committee to develop a full description of the ACE method, but has yet to issue a report; (d) a recent document from the IAI to the National Academy of Sciences stated that they needed to produce agreement on the description of the method (Martin, 2007); and (e) the Office of the Inspector General of the Department of Justice (Fine, 2006), the internal analyses of the FBI's erroneous identification of the Madrid bomber (Budowle, et al., 2006), and the National Academy of Sciences Report on Forensic Sciences (2009) each indicated that the method was not spelled out completely.

Fifth Basis: The Profession Has No Official or Approved Description of the ACE Method.

The fingerprint profession has never constructed or approved an official description of the ACE method to be used to compare fingerprints. No document issued by the National Institute of Justice (NIJ), the International Association for Identification (IAI), the Technical Working Group on Friction Ridge Analysis, Study and Technology (TWGFAST), the Scientific Working Group for Friction Ridge Analysis, Study, and Technology (SWGFAST), any training program authorized by the IAI or the American Society of Crime Laboratory Directors (ASCLD), or any single crime laboratory, including the FBI, has ever stated that it has an approved version of the ACE method. There is no community made up of fingerprint examiners that agrees to the description, the steps, and the standards of the ACE method. Until the fingerprint profession defines a particular ACE method, there is no particular version for the fingerprint community to accept.

Sixth Basis: Published Descriptions of the ACE Method Differ from One Another.

The most widely respected texts used by examiners, including Ashbaugh (1999), Champod, et al., (2004), Conger (1993), Olsen (1978), and the FBI (1988), differ from one another in the description of the ACE method, including critical steps in the method and their sequence. Several recently published comparisons by Beeton (2002), Triplett & Cooney (2006), and by Haber & Haber (2008; 2009, chapter 5) point out the significant disagreements among the authors. As two examples, we describe two different feature descriptions and three different relative location specification systems used in current practice during the analysis stage of ACE.

Feature Descriptions. Most, but not all examiners in the United States use the Galton (1896) minutiae system as the basis for fingerprint comparison. Recently, some examiners use a newer feature description, called quantitative-qualitative friction ridge analysis (Ashbaugh, 1999) or ridges-in-sequence (Ashbaugh, 2005). To date, the profession has not determined which procedure to describe the features present in a latent constitutes application of the ACE method.

Spatial Location Specification. The specification of the exact location of each feature (e.g., ridge ending, bifurcation), in relation to every other feature, constitutes a fundamental descriptive requirement for analysis of the latent and subsequent comparison to the exemplar (Stoney, 2001), whichever feature system is applied. Three different spatial location specification systems are employed at present: subjective eyeballing, ridge count, and ridges-in-sequence.

Subjective eyeballing, in which no formal measurement is performed, appears to predominate, based on our experience from hearing examiners' testimony in court. Because the location of the features in the latent can never exactly match their location in a second deposition from the same finger, the examiner makes a subjective decision for each correspondence: These two features look as if they are in about the same location in the latent and exemplar.

Ridge count is a somewhat more objective measure. The examiner counts the number of ridges that intervene between two Galton features. Spatial location agreement depends on the same ridge count in the exemplar as in the latent. This method has two drawbacks. It does not specify the location of features in any absolute measure, it only counts the number of intervening ridges in each print. Further, the flexibility of the skin often makes ridge counts indeterminate, because the number of ridges in any

portion of a latent print partly depends on deposition pressure. This means that intervening ridge counts between corresponding features can differ in a single donor latent and exemplar comparison.

Ridges-in-sequence is a new spatial position system, in which the examiner specifies the locations of features in relation to one another along a single ridge path, and then along each adjacent ridge path. The system remains undefined: should the curvature of the ridge be measured? Should the angle of the arms of a bifurcation be specified? Should the distance between two features be measured? At present, ridges-in-sequence is another eyeballing procedure, one that differs from the purely subjective older version in that the feature locations are specified in sequence along individual ridge paths.

To date, the profession has not determined which feature description and which spatial location procedure constitutes application of the ACE method.

Seventh Basis: Training, Proficiency, and Certification Programs are Not Linked to a Specified Comparison Method.

The absence of an official description of the ACE method means there is no syllabus on which to base training of the ACE method. Consequently, different training programs use different training procedures. There are no standardized tests used at different stages of training to document mastery of the components of the method. There is no standardized final examination to be passed at the end of training, and there is no specific ACE method on which to base subsequent proficiency and certification testing.

Eighth Basis: Examiners are Not Required to Make Contemporaneous Bench Notes When Performing a Comparison.

There is no professional requirement that examiners in this country document the steps taken and the basis for conclusions reached in a comparison. We have never seen an examiner provide contemporaneous bench notes in court. No proficiency test and no experiment reportedly assessing examiner accuracy has required bench notes to show ACE was used. Consequently, there is no evidence that examiners followed the ACE method, or that they applied it correctly.

In summary of the professional acceptance of ACE, research scientists do not accept ACE, and so testify on the grounds that it lacks evidence of validity and reliability. The evidence that practitioners do not accept ACE is less direct, but equally compelling. The profession has not defined ACE as a method beyond a general framework, they have been unable to agree among themselves about the steps, and they do not demand any evidence that examiners use the ACE method in their work, their proficiency testing, their certification testing, or even in the few experiments conducted about the accuracy of examiners.

Opinion 2: No Scientific Studies and Reviews Are Published in Peer Reviewed Scientific Journals Supporting the Scientific Validity of ACE or of Fingerprint Comparisons.

A second criterion applied under Daubert is whether a sustained publication history of research in peer-reviewed scientific journals supports the validity of the ACE method.

Haber & Haber (2009, chapter 8) reviewed the entire body of peer-reviewed published research on the results of fingerprint comparisons prepared by research scientists. Their search covered the 130 years from the beginning of modern forensic fingerprint comparison work to the present. We list all of the experimental work published that presented new data, whether the experiment controls for the use of the ACE method or not (citations on reference list).

Anthonioz, Champod et al, 2008

Black, 2006

Cole, Welling, Dioi-Vlla and Carpenter 2008

Dror & Charlton, 2006

Dror, Charlton & Peron, 2006

Evetts & Williams 1996

Hall & Player 2008

Langenburg 2009

Langenburg 2004

Langenburg, Champod and Wertheim 2009

Wertheim, Langenburg & Moenssens 2006

Tangen & Vokey 2009

Vokey, Tangen & Cole, 2008

First Basis: The list contained in Opinion 2 in Section B does not meet a Daubert criterion of a substantial body of publications.

Thirteen articles in 130 years is a tiny number.

Second Basis: Not one of these articles contains any evidence of the accuracy of the ACE method.

Examiners in these studies were not required to show how and why they arrived at a conclusion: there is no evidence that they used ACE, or used it correctly. Nearly all of these articles are reviewed elsewhere in this affidavit, because their results also suggest poor validity of whatever comparison method was used.

Opinion 3: The ACE Method Cannot Be Tested for its Validity.

First Basis: Until ACE is Fully Described and Officially Adopted by the Profession, there is No Single ACE to Test.

The fingerprint profession has neither specified nor approved a version of the ACE (see Opinion 1 in Section B, bases 4-8 above). The NIJ, IAI, TWGFAST, SWGFAST, ASCLD have not endorsed any particular version of ACE.

There are nearly a dozen textbooks and manuals written by respected members of the fingerprint profession. They differ from one another in their description of the method. As one example, Ashbaugh (1999) requires complete analysis of the latent before the examiner views the exemplar; many texts, including the FBI's Science of Fingerprints (1988) do not specify this step, or they permit initial simultaneous side-by-side comparison. There is no agreement as to how to define spatial location, or how to determine when two features are in agreement in spatial location. At present, ACE refers to some methodological steps that have not been well described, and to differing procedures (Champod, 2008; Cole, 2008; Mnookin, 2008). Until the method is specified and endorsed, there is no method to test. Until there is a standardized, approved description of the method, the method cannot be validated. Which method should the validation experiment test?

Second Basis: ACE Cannot be Validated Because there is No Objective Measure of Latent Print Information Content.

A validity test of ACE requires latent-exemplar pairs that correspond in difficulty to those encountered in casework. There is no metric for the difficulty of a latent (see Section C, Opinion 1). Until this metric is developed, ACE cannot be validated, because there is no objective measurement of the latent on which to base a value standard. There can be no assurance that the latents used in a validation experiment are equivalent to those found in casework.

Third Basis: ACE Cannot be Validated Because the Proportions of Conclusions in Casework is Unknown.

A validity test of ACE requires the same proportion of latents of value, exclusion, identification and inconclusive as occur in casework. What percent of casework latents are of value? What percent exclusions? The profession maintains no record of the proportion of these conclusions. At present, only anecdotal percents are available (Haber & Haber, 2009, chapter 8). Until these proportions are known, the latents in the validity experiment cannot be demonstrated to represent casework and the results from the experiment may not represent accuracy of ACE outcomes in casework.

Fourth Basis: Training Programs and Proficiency Testing are Not Designed Around an ACE Method.

A validity test of the ACE method requires examiners as subjects who are highly proficient in the ACE method. Proficient examiners are necessary to rule out poor examiner skill in computing the error rate of the method. Examiners cannot document their proficiency in using the ACE method, because the profession lacks validated, standardized training programs built around the ACE method that include validated and reliable proficiency tests. At present, proficiency tests are profoundly inadequate (see Section E, Opinion 2, and Haber & Haber, 2009, chapter 7) and do not require evidence that the examiner is applying ACE during the testing. Until adequate proficiency tests are developed, including documentation that the examiner is highly proficient in using ACE to reach his conclusions, ACE cannot be tested for its validity.

Fifth Basis: ACE Cannot be Validated Because Examiners do Not Make Contemporaneous Bench Notes of their Application of ACE.

A validity test of ACE requires evidence that the examiners are applying ACE correctly. The bench notes must contain specified information at each step, in sequence. This uniform report form needs to be written and adopted by the profession. At present, examiners using the ACE method in this country are not required by anyone to record the steps they took to reach their conclusion and no standardized report form has been tested and approved. Until examiners document their work in detail, in standardized form, ACE cannot be tested for its validity. No published experiment asked the subjects to record bench notes regarding the steps they used to arrive at a conclusion.

Opinion 4: There is no Scientific Evidence of the Validity and Reliability of the ACE Method: the Method Has Never Been Tested.

The Validity of a method is concerned with the correctness of the conclusions reached. A fingerprint examiner attests to an identification in court. She testifies as a scientific expert who offers her opinion on the basis of application of a scientific method called ACE. Is the testimony accurate? For a fingerprint examination, when an examiner concludes identification and there is certain knowledge (called ground truth) that the exemplar and latent were made by the same finger, the method has produced a valid (correct) conclusion. Similarly, when an examiner concludes exclusion, and there is certain knowledge that the exemplar and the latent were made by different fingers, the method has produced a valid conclusion. Validity of the ACE method is demonstrated when an experiment shows that examiners employing the method are able to reach correct conclusions.

The Reliability of a method concerns consistency, and not accuracy. If highly trained and experienced examiners applied the ACE method to compare a large number of latent-exemplar pairs and they all reached the same conclusion about each pair, ACE would be shown to be a reliable method. Also, if the same examiners, years later, compared the same latent-exemplar pairs using the ACE method and each examiner reached the same conclusion about each pair as before, ACE would be shown to be reliable. Reliability refers to consistency among the examiners applying the method: they agree with one another, and they make the same judgments over time.

A reliable method is not necessarily a valid method. Suppose these examiners all reached the same but wrong conclusions. The method would be reliable, but invalid. Most knowledgeable people used their vision to opine that the world was flat (a reliable conclusion), though they were wrong (vision alone can lead to invalid conclusions). If a method is unreliable, so that different examiners reach different conclusions about the same latent-exemplar pair, the method is necessarily invalid, because some of those examiners are reaching the wrong conclusion. We define the scientific distinction between reliability and validity because the Daubert rulings refer to reliability as inclusive of accuracy. In this affidavit, we always use the term validity to refer to accuracy (agreement with ground truth), and the term reliability to refer to consistency or agreement among examiners.

A pure test of the accuracy (validity) of the ACE method requires a careful experiment. The ACE method must be precisely described, and skilled examiners serve as subjects to carry out the comparisons, recording their work so it can be determined that they are applying ACE properly. The examiners must be selected based on high levels of training, experience and demonstrated proficiency. The test materials must consist of a wide range of latent fingerprints typical of the kinds of prints that are found in casework. In the experiment, each latent print is paired with an exemplar print for which ground truth of the pairing is known. The experiment is conducted under optimal working conditions. The results are scored only for those responses for which ACE was properly applied. Scoring reflects the number of correct and incorrect conclusions. With these specifications, the number of correct conclusions reached represents the accuracy of the ACE method, and the number of erroneous conclusions made represents the error rate of the method, independent of errors made by examiners.

First Basis: There is No Published Experiment of this Kind.

Budowle, et al. (2006), Cole (2005), Haber & Haber (2004), Haber & Haber (2008), NAS (2009) and Haber & Haber (2009, chapter 2) searched in vain for a published experiment demonstrating the validity of the ACE method under these conditions. The necessary research to demonstrate the error rate of the ACE method has never been performed.

Second Basis: Judicial notice has been Given of the Lack of Validity of ACE.

In a Frye hearing prior to the trial of Maryland v. Bryan Rose (2007), Judge Souder ruled that fingerprint evidence was not admissible because it lacked evidence of its validity and reliability (Souder, 2007; 2008).

Third Basis: Research Scientists Assert that the Validity of the ACE Methodology is Unknown.

When research scientists have offered testimony or signed amicus briefs in these 50+ hearings, their opinions have strongly voiced support for a motion to exclude fingerprint evidence on the grounds of the lack of scientific evidence of its validity.

Fourth Basis: The National Academy of Sciences Criticized the Absence of Scientific Evidence for the Validity of ACE.

A recent report from the National Academy of Sciences on forensic science provided a thorough review of the scientific support for fingerprint comparison. That report commented on a number of inadequacies in evidence in support of ACE, including the

absence of evidence of its validity: “ACE provides a broadly stated framework for conducting friction ridge analyses. However, this framework is not specific enough to qualify as a validated method for this type of analysis” (NAS, 2009, 5-12).

Opinion 5: Because the Validity of ACE has not been Tested, the Error Rate of the ACE Method is Unknown.

First Basis: The only proof of an error rate of a scientific method requires an experimental demonstration using fingerprints of known ground truth, examiners of known proficiency, and evidence that the method in question was applied correctly. Since these conditions have not been met for the ACE method, there is no known error rate for ACE.

Opinion 6: Non-empirical Grounds Cannot Substitute for a Scientific Experiment to Show the Error Rate of ACE and do not Suggest a Known Error Rate.

Fingerprint examiners have testified as proponents of the ACE method in admissibility hearings that non-experimental reasons indicate the error rate of the ACE method is (virtually) zero. It is our opinion that none of these non-experimental measures indicate a known error rate for ACE. The five bases for this opinion rest on five different claims made by the fingerprint profession.

First Basis: History of Acceptance is Not a Measure of Validity.

The 100-year-old historical acceptance of fingerprint evidence by the courts is claimed to indicate that the method's error rate is very low or zero. This argument has no scientific basis. Validity is a measurement of the accuracy of the outcome of the application of a method. A court's acceptance does not make a method valid, only accepted.

Second Basis: Certainty Does Not Make Conclusions Valid.

Fingerprint examiners claim that the error rate of the ACE method is zero, because they only testify to a conclusion when they feel absolutely certain it is correct. Certainty is not well correlated with accuracy. To the contrary, scientific research has repeatedly shown that certainty is not a measure of correctness, and that the correlation between certainty and accuracy is very low under normal circumstances (Wells, Menon & Penrod, 2007). Further, the fingerprint profession has required examiners to testify with “100% certainty” (SWGFAST, 2004). The assertion of certainty, given this requirement, reflects to an unknown extent the requirement itself, not to some evidentiary base.

The National Academy of Sciences report (2009), quoting J.L. Mnookin, criticized this inflated testimony: “Given the general lack of validity testing for fingerprinting; the relative dearth of difficult proficiency tests; the lack of a statistically valid method of fingerprinting; and the lack of validated standards for declaring a match, such claims of absolute, certain confidence in identification are unjustified” (NAS, 2009, 5-12).

The assertion of 100% certainty has received judicial notice. In *USA v. Plaza I* (2002), Judge Pollak ruled that an examiner could describe to the jury how he performed the comparisons, and could show the jury a marked up version of the exemplar and the latent prints being compared, but he could not offer any opinion to the jury as to whether the two prints originated from one or two donors, nor provide any indication of confidence in any possible opinion. Judge Pollak in *USA v. Plaza II* (2002) reversed his own ruling. Cole (2004) provides a detailed analysis of both rulings by Judge Pollak.

Third Basis: The Claim that “the Error Rate of the Method is Zero Because All Errors Can be Assigned to Examiner Error” is Factually Untrue.

The fingerprint profession has claimed that the error rate of the ACV method is zero, and that therefore, when errors are made, they are always due to inexperience or poor training of examiners (Meagher, 2003; United States v. Havvard, 2000). By this tautology, the method itself is perfectly valid and error-free.

A tautology is not scientific evidence, it is only a circular and illogical argument. In this case, the premise of this tautology is also known to be factually wrong. The recent FBI erroneous identification of Brandon Mayfield was made by three of their most experienced and skilled examiners, and verified by an equally skilled examiner independent of the FBI (Stacey, 2004; Smrz, et al., 2006). Most of the other documented erroneous identifications made in court were contained in testimony of highly experienced examiners, often certified ones (Cole, 2005).

Fourth Basis: Verification Does Not Necessarily Reduce the Error Rate to Near-zero.

The fingerprint profession argues that whenever the conclusion offered in court has been verified by another examiner, its validity must be very high. However, research evidence shows that non-blind evaluation of performance is open to significant bias, and, compared to blind verification procedures, has a much higher failure rate of catching errors (Arvizu, 2002; Boone, et al., 1982; Food & Drug Administration, 2001). Non-blind verification is no more than a ratification of the first examiner's work (Haber, 2002), not an independent (blind) replication. The FBI acknowledges that their verifications are made without blind procedures (Budowle, et al., 2006). There are no requirements in place from the profession (e.g, SWGFAST, 2002b) that blind verification is mandatory. Further, many of the erroneous identifications attested to in court were verified (Cole, 2005).

More seriously, casework examinations and verifications are made without knowledge of ground truth. No one has absolute knowledge whether the two prints being compared were made by one or two donors. Therefore, both examiner and verifier each could be using a method that can reach a wrong conclusion for the particular fingerprints involved, and never discover that fact, because the prints came from casework without known ground truth (Mnookin, 2008).

Fifth Basis: Adversarial Argument Does Not Reduce the Error Rate to Near-zero.

The fingerprint profession has argued that in the adversarial process of a criminal trial, the validity of conclusions will be uncovered and made evident to the jury (United States v. Havvard, 2000). The profession claims that because an erroneous identification has never been uncovered during the course of an adversarial proceeding, no erroneous conclusions based on the ACE method have been offered in court (Cole, 2005). This argument has no relevance to the validity of the ACE method. In court, ground truth is unknown; there is no evidence that the adversarial process uncovers errors; there is no evidence that the adversarial process uncovers errors any specific percent of the time; there is no evidence that the adversarial process is the most accurate procedure to uncover errors; and there is no evidence that the examiner used the ACE method to reach the conclusion. We discuss (see Section B, Opinion 2) the existence of over 30 cases of erroneous identifications that survived the adversarial process without detection.

C. OPINIONS REGARDING FACTORS THAT COMPROMISE THE VALIDITY OF THE ACE METHOD.

We have already opined (Section B, Opinion 4) that no published research studies exist on the validation of the ACE method. In this section, we consider four important gaps in the description of ACE that implicate further poor validity. These include: subjective measurements of fingerprints during comparison; subjective standards used to control conclusions made during comparison; incorrect scoring of conclusions in case work, proficiency tests and in experiments; and validity problems with the use of automatic search systems to produce candidates for comparison with ACE.

Opinion 1: The Measurements Made on the Fingerprints being Examined in ACE Are Subjective.

Each of the three stages of ACE requires an objective measurement of the physical evidence found at that stage: (1) in Analysis, the information content of the latent, (2) in Comparison, the kind of dissimilarity found between latent and exemplar, and in Evaluation, (3) the amount of similarity found between latent and exemplar. It is our opinion that none of these physical measurements is objective. We describe these measures for each stage of ACE.

Analysis Stage. The examiner determines whether the latent print contains sufficient quantity and quality of information to use for further comparison (he makes a decision of *value*). To oversimplify, quantity refers to the size of the print and the amount of feature detail it contains. Quality refers to the clarity of the print, whether the detail is clear or ambiguous. How much quantity and how much quality, in what relation to one another, are needed for an examiner to use the print for a comparison? Among the components of quantity and quality suggested by the profession are: the size of the print, the number of features clearly visible, whether orientation is obvious, amount of contrast, amount of distortion, amount of continuous detail (in contrast to prints taken from rumpley surfaces such as plastic). Each of these components is physically measurable (see Ashbaugh, 1999, for a detailed description). For each dimension, an objective scale could be established. An experiment could be performed to determine how to weight these components relative to one another (clarity might be more important than size, for example). The results would establish an overall metric for the amount of information in a latent. At present, there is no metric in ACE for the amount of information the latent contains. Examiner judgment about the amount of information in a latent is subjective.

Comparison Stage. Since every take of a fingerprint is different, even from the same finger, a latent and exemplar print will always be different. The examiner has to distinguish two causes of those differences: did they arise from distortions created by the manner of the deposition, by the surface on which the print was left, and from the processing of the print (a conclusion suggesting a possible identification); or did they arise from discrepancies between the two prints because there were made by two donors (which would trigger an exclusion conclusion). This discrimination requires an objective measurement of the kinds and amounts of differences found when multiple takes of the same finger are compared.

While some textbooks (e.g., Ashbaugh, 1999) discuss distortions, none quantifies distortion. To oversimplify, as an example, research could establish the changes in the shapes of ridges and grooves resulting from different deposition pressures: they are entirely predictable and could be described physically in terms of the direction and amount of pressure exerted. At present, there is neither a description nor a metric in ACE for the kinds of predictable distortions that occur in a latent. Nor has there been any research on whether examiners can differentiate among sources of distortions, or distinguish which differences are due to discrepancies. Without objective measurement of the differences and their causes, examiner judgment about the presence of distortion in the latent and discrepancy between a latent and an exemplar is subjective.

Evaluation Stage. The examiner determines the amount of similarity between the latent and exemplar. Amount of similarity must be quantitative and measurable, once the features to be compared and similarity are objectively described. At present, there is no metric in ACE for the amount of similarity between two prints, so that no objective measure of the amount of similarity between two fingerprints is presently in use by the fingerprint profession.

Prior to 1973, examiners in the United States were trained to count the number of minutiae or points they had found in the two fingerprints that were in corresponding locations. This count was used as a measure of similarity. The greater the number of points in agreement by kind and location, the greater the similarity between the two prints. Assuming that examiners could find features, locate them on the fingerprint using a metric of spatial location, and could count the number of these accurately, then the number of points-in-agreement was an objective measurement of similarity.

However, in 1973 the IAI abandoned the point count measure when they could find no evidence to support it as useful for an objective measurement of similarity (this position was reiterated by the IAI (IDentification News, 2009, vol. 39, No. 6, p. 20).

The fingerprint profession has not replaced the point count measure with another objective measure of similarity, though most examiners testify at trial that they still use some variation of disallowed point counting for a measure of similarity. Without a

measure of similarity, examiner judgment about the amount of similarity between two prints is subjective: individual examiners make their own judgment of the amount similarity between two prints.

The three measurements of the physical evidence in a latent print or similarity between two prints have no objective basis: they are entirely subjective. Until objective measurements for these are developed, taught, and validated, the ACE method itself cannot produce validated results.

Opinion 2: The Threshold Standard for Conclusions at Each Stage of ACE is Subjective.

Unlike judgments of (1) amount of information content, (2) type and amount of distortion, and (3) amount of similarity between two prints, all of which should be objective attributes of the target prints to be compared, standards are arbitrary objective thresholds established by the profession. For example, age is objectively defined as elapsed time since birth. Applied to this objective age measurement are several arbitrary standards: it is used to regulate the right to drive, to fight for your country, to vote and to receive social security. Each of these arbitrary thresholds is determined by the governing body and applied to that objective measurement of age to reach a conclusion of the law.

Conclusions based on the ACE method of fingerprint comparison are expected to differ between examiners because, unlike an age threshold, the standards themselves are subjective (Haber & Haber, 2007; 2009, chapter 5). The subjective nature of each standard implies that if an experiment were performed to assess the validity of the standard, a high error rate would be revealed. However, no experiments have evaluated the application of any of these standards against ground truth. In the absence of such tests, each standard lacks validation, and conclusions based on these standards also lack scientific validity.

Analysis Stage. A fingerprint examiner must decide, as part of the Analysis Stage of ACE, whether the information content of a latent print meets a value standard for further comparison. Without an objective measurement of the content of a latent fingerprint, the profession cannot define this standard. The standard could be based on research evidence showing that latents of information level two (not much quantity and quality) lead to erroneous conclusions 15% of the time, whereas latents of information level four lead to erroneous conclusions only 5% of the time. If the profession accepts 5% as the maximum amount of error it will tolerate, the profession would set the value standard at latent information level four. This research has never been performed, and there is no standard for value that is approved by the profession. Examiners' conclusions about the value of a latent are subjective. Subjective judgments lead to variation in judgments among examiners: different examiners make different value decisions about the same print. When decisions about the same print differ, the conclusions are unreliable and invalid: some of them are wrong. Smith (2004) showed substantial variation in the value/no value decision in working examiners.

Comparison Stage. The exclusion standard is used to determine whether the differences between the latent and the exemplar are due to distortions created when the same finger made the two fingerprints (one donor, so continue ACE), or due to discrepancies that must be the result of two different fingers making the two prints (two donors, so exclude the exemplar as the source). The profession has an objective standard, sometimes described as the "one inexplicable discrepancy" rule (SWGFAST, 2002). The standard says that even if only one discrepancy is found, the examiner must exclude the suspect as the contributor of the latent print.

However, because the underlying physical measurement of discrepancy resulting from two donors is not objectively defined, examiners discount a single discrepancy or ignore discrepancies until they have found many of them. From our experience, many examiners do not know about the one inexplicable discrepancy exclusion standard, or do not know its logical derivation from the uniqueness assumption. As an example of the latter, one prominent examiner (Leo, 1998) discusses this distinction and asserts that if the examiner has found a sufficient number of matching characteristics, two-donor pattern differences cannot occur. Even with an objective standard, the conclusions reached remain subjective.

Evaluation Stage. The sufficiency standard is used by an examiner to decide whether there is sufficient similarity between a latent and exemplar to conclude that the same finger made both fingerprints. The fingerprint profession relies on an entirely subjective judgment: each examiner decides what is sufficient based on his own experience and training.

If there were an objective measurement of similarity (which might be the number of points or adjacent ridges in agreement), then a simple experiment could determine the threshold standard for an identification. Similarity would be measured in a large number of latent-exemplar pairs of known ground truth, and that similarity score compared to whether the pair was from one or two donors. If 95% of all pairs with four or more adjacent ridges in agreement were one donor pairs, then using four or more matching adjacent ridges as the threshold standard would limit the error rate for this conclusion to 5%. If the profession wanted to limit the error rate to 2%, a higher number of adjacent ridges in agreement would be required before an identification conclusion could be reached.

This procedure, based on research, makes the similarity standard objective, determined by a given error rate. The profession would decide what probability of error it accepts and set the sufficiency standard accordingly. This research has never been performed. Examiners' conclusions to identify are subjective. Subjective judgments lead to variation in judgments among examiners: different examiners reach different conclusions about the same latent-exemplar pairs. When decisions about the same prints differ, the conclusions are unreliable and invalid: some of them are wrong. The experimental results from Evett and Williams (1996) document the wide variability in conclusions reached by skilled examiners.

In *USA v. Parks* (1991), Judge Letts did not allow the fingerprint evidence to be introduced into his federal court because examiners disagreed on the standards for sufficiency of agreement necessary to conclude identification. Cole (2001) has a detailed discussion of the importance of Judge Letts' decision.

Opinion 3: The Fingerprint Profession Incorrectly Computes the Error Rate of ACE.

The fingerprint profession claims that only the erroneous identification rate is relevant to the calculation of the error rate of ACE, because only erroneous identifications have serious consequences. While we agree that erroneous identifications are heinous, scoring only erroneous identifications distorts the assessment of the validity and error rate of ACE. In our opinion, the correct computation of the validity of ACE must include each decision the examiner makes when ACE is applied.

When a fingerprint examiner applies ACE, he opines only one of four conclusions (no value, exclusion, identification, or inconclusive). Comparing those to the two possible ground truth values (same donor, different donors) shows there are two possibilities for correct judgments and six for erroneous judgments. These are shown here, with the correct judgments in boldface.

TABLE

We offer four bases for our opinion.

First Basis: The Validity of the ACE Method is Its Accuracy in Discriminating Between One and two Donors.

The fingerprint examiner's task, applying the ACE method, is to compare two prints and determine whether they were made by the same or different fingers. The method's validity is the accuracy with which it makes this discrimination when properly applied. Therefore, the correct terms to use when computing the validity of ACE are: the number of correct identifications and correct exclusions combined (the numerator) divided by the total number of comparisons performed (the denominator). We have never seen this computation applied in any experiment, or to any proficiency or certification test results.

Second Basis. The power of ACE to Discriminate Between One and Two Donors is Important to Society and to the Judicial System.

Whenever the application of ACE by a skilled examiner results in an erroneous exclusion or an inconclusive judgment, either an innocent person remains at risk of indictment or a guilty person may remain at large. Contrary to the fingerprint profession's claim that only erroneous identifications are relevant to the court, the frequencies of erroneous exclusions and inconclusive judgments are relevant to society and to the judicial system.

Third Basis. Examiners are Encouraged to Exercise Extreme Caution in Making Identifications, so that Failure to Consider all Six Cells Over-represents the Validity of ACE.

The fingerprint profession severely disciplines an examiner who makes an erroneous identification and encourages examiners to be very cautious in making identifications. As a consequence, examiners make fewer identifications. This necessarily increases the number of inconclusive judgments and erroneous exclusions (see our next basis). When inconclusive judgments and erroneous exclusions are not scored, and only erroneous identifications included, the validity of ACE is artificially inflated.

Fourth Basis: The Fingerprint Profession's Focus on Erroneous Identifications Prevents Assessment of the Validity of ACE.

When the only conclusion considered is identification, no data are available to study the impact of changes in the relative frequency of the different kinds of conclusions that are entirely under the examiner's control. If an examiner wishes to reduce his erroneous identification rate, he can easily do so by making fewer identifications, and shifting some of what would have been identifications to insufficient or exclusion conclusions. The result is that the examiner reduces his erroneous identification rate, thereby appearing to be a more proficient examiner and making ACE appear to be a more valid test. This shift reflects only a change in confidence, or a willingness to make a risky conclusion, but does not reflect any improvement in the power of the ACE method to tell the difference between one and two donors--the validity of ACE (Swets, Dawes & Monahan, 2002). Evidence of the power of the method to discriminate between one and two donors is not altered by a change in the willingness or confidence of an examiner to make identification conclusions (Swets et al., 2002).

Opinion 4: Biases Introduced in the Use of Automated Search System (AFIS) Procedures Increase the Probability of ACE Erroneous Identifications.

Two aspects of the use of AFIS to produce suspect exemplar fingerprints have the potential to increase the error rate of the subsequent examination of those exemplars by the use of the ACE method.

First Basis: Preprocessing of Latent Prints Introduces Potential Bias in the Application of ACE.

Most AFIS systems require either that a technician or examiner preprocess the latent before submission for search. Whenever this occurs, changes are made in the latent that alter its information. Each change re-interprets, with unknown accuracy, details in the latent. The processed latent no longer matches the original latent, in that ambiguous features have been disambiguated. This may be useful to the AFIS system, but it has the potential to increase the ACE error rate.

Preprocessing increases bias into the ACE comparison method if the examiner performing the comparison was the person who did the preprocessing, or saw the preprocessed latent afterwards, or used the preprocessed latent during the comparison to the exemplars found by AFIS. Research results from a range of studies have shown that once you "know" the image, those details remain clear and you remain certain they are accurate (Risinger, et al., 2002). As a consequence, the examiner is much more likely to conclude that the latent matches the AFIS exemplar. The outcome of the comparison is biased in two ways. First, the

examiner “sees” the features he interpreted as present. Second, the AFIS output is biased by algorithm to produce candidates who possess precisely those features. Both biases increase the probability of an erroneous identification.

Second Basis: Biases Associated with Cold Hits.

When an examiner individualizes a latent to an AFIS candidate who is not associated with the crime in any other way, that suspect is called a “cold hit.” If subsequent investigation fails to produce other evidence of complicity in the crime, this suspect remains a cold hit. A hot hit, in contrast, occurs when the fingerprint examiner identifies a person who already was suspected as the perpetrator based on other evidence.

According to statistical analyses based on Bayes Theorem, hot hits and cold hits from AFIS have entirely different statistical likelihoods of accuracy: cold hits have a vastly higher probability of being erroneous identifications than hot hits made by the same examiner.

Consider a fingerprint examiner with a tiny erroneous identification rate of 1% (one erroneous identification in every 100 identifications he reports). Assume also that chances are 1 in a 1,000 that the man identified by the cold hit is the perpetrator. From these data, laypersons, the fingerprint profession, and the courts treat the chances of error as 1%--the error rate for the examiner. In fact, the odds of an erroneous identification are computed as a combination of the examiner error rate and the probability that the cold hit committed the crime. In this example, the .01 error rate for the examiner, when combined with the .001 probability that the identification is correct, results in a true accuracy of only 10%, not 99%. There is a 90% chance that the identification is erroneous, in spite of the examiner's normal high accuracy rate (Thompson, et al., 2003).

AFIS cold hits involve another bias problem. Once the fingerprint examiner makes the identification of a cold hit, the police usually seek evidence to confirm it, while often ignoring disconfirming evidence. A wide research literature demonstrates that confirmatory bias leads to erroneous conclusions (Risinger, et al. 2002).

D. OPINIONS REGARDING EVIDENCE SUGGESTING A LACK OF ACCURACY FROM FINGERPRINT COMPARISONS.

Opinion 1: Research Evidence Shows that the ACE Method Is Not Applied Reliably or Accurately by Fingerprint Examiners.

Nearly all of the fourteen empirical studies listed in Section B, Opinion 2 have reported disagreement among skilled examiners during the Analysis, Comparison, and Evaluation stages of ACE. When inconsistencies are found among highly trained and experienced examiners, they point to failures in the method, not the examiners. If examiners reach different conclusions about the same latents or latent-exemplar pairs, some are wrong and the method employed is neither valid nor reliable. We summarize results from nine of the fourteen studies here, subdivided by the stage of ACE being tested. In combination, these experimental results indicate that whatever the method used by fingerprint examiners has poor validity.

Analysis stage: Smith's Value / No Value Tests

Smith (2004) tested examiners' ability to distinguish between latent prints of value and those of no value, using easy latents. The results showed that working examiners made about 15% errors on a value/no value conclusion for relatively easy latent prints.

Analysis stage: Langenburg's Test of Galton Feature Labeling

Langenburg (2004) asked experienced examiners to examine 14 latent fingerprints and mark down the number of Galton points they found in each latent. His results showed that examiners extensively disagreed whether the Galton features are present or absent in latent prints, and on how many features were found.

Analysis stage: Black's Simultaneous Latent Print Study

A simultaneous latent is defined as an impression containing two or more prints from the same donor made at the same instant in time. Black (2006) reported a study to determine whether latent fingerprint examiners can successfully discriminate between simultaneous and non-simultaneous latent prints. Black asked 31 trained examiners to examine 30 latent multiple impressions and to conclude, for each one, true simultaneous, false (non-simultaneous), or inconclusive if he couldn't tell.

Of the 930 responses, 252 were inconclusive (27%) and 83 were erroneous (9%), or 36%. In over 1/3 of their analyses, these examiners could not successfully determine whether the prints were simultaneous.

Comparison Stage: Evett & Williams' Study of Agreement of Galton Features

One hundred thirty highly experienced examiners compared ten latent-exemplar pairs; all of the latents were of value for comparison, and nine of the ten pairs were identifications. Permitted responses were identification, probable identification (an English response category indicating 8 to 15 points of agreement), insufficient information, or exclusion. No erroneous identifications were made. There were ten erroneous exclusions (0.8%). It is not known which version of ACE was being used, if any, nor was this research designed or intended to show examiner accuracy rates.

One of the important findings from the study concerned the low reliability of reporting the similarity between the two fingerprints in each pair. The authors found substantial disagreement among the 130 examiners as to the number of Galton points in correspondence in each latent-exemplar pair. For one pair, the 130 examiners gave responses that ranged from 14 to 56 features in agreement.

Comparison Stage: Langenburg, Champod & Wertheim (2009)

Forty-three skilled examiners reported the number of features they found in agreement in 3 same-source latent-exemplar pairs. As in the Evett and Williams study, a wide range of reported corresponding minutiae was found, indicating a high level of unreliability. Overall, the smallest difference between examiners was 8, the largest difference 21.

Evaluation stage: Evett and Williams (1996) Conclusion Disagreements

The examiners in the study described above were also asked to record their conclusion for each of the ten latent-exemplar pairs. The examiners differed in their conclusions for all ten pairs. For one pair, more than half of the examiners (54%) concluded a probable identification, and the remaining examiners were divided between insufficient detail (38%) and exclusion (8%). This print was judged by some examiners to be an identification (probable with 15 points of agreement), and by others an exclusion (the exemplar was from a different donor than the latent).

Evaluation stage: FBI's Mitchell Survey Results

Byron Mitchell was tried in federal court for allegedly driving a get-away car in a robbery. Two latent fingerprints were found on the gear shift lever, which an FBI examiner identified to Mr. Mitchell's exemplar prints. When the defense challenged the FBI's lack of evidence for fingerprint identification accuracy, the FBI offered a demonstration (reported in *US v. Plaza I*, 2001). They mailed a packet containing Mr. Mitchell's ten print card along with the two latent prints found in the car to 50 different crime laboratories in the US and asked each laboratory to have its most experienced examiner compare these two latents to the exemplar and indicate his conclusions. Thirty-nine laboratories responded. Thirty reported that they identified both latents to Mr. Mitchell, 9 identified only one or neither to Mr. Mitchell. Twenty three percent of these experienced examiners disagreed with the FBI conclusion.

Evaluation stage: Two Experiments by Dror, et al. Show Bias Leads to Changed Conclusions on Repeated Examinations

Two experiments using highly skilled working examiners show that extraneous information influences examiners, and increases the number of erroneous conclusions they reach.

Dror and Charlton (2006), using pre-selected examiners, obtained from each of their past casework a latent-exemplar pair that the examiner had concluded to be an identification at least five years previously. When this same pair was presented again (none of the examiners recognized it as familiar), each examiner was told that the latent was from a notorious case of an erroneous identification. Over two-thirds of the examiners now concluded either exclusion or uncertain. Only 20% reiterated their initial (presumably correct) conclusion of identification.

In a second experiment, Dror, Charlton and Peron (2006) used a larger sample of examiners and latent-exemplar pairs. From each examiner's past casework, they picked pairs in which half of the pairs had been identified and half excluded. The pairs were presented anew to the same examiners who had compared them previously (none recognized the pairs as familiar), but with extraneous information that served to confirm or contradict the examiner's former conclusion. The information given influenced the examiners' current conclusions in predictable ways. The examiners who received information inconsistent with their previous conclusion were more likely to reach a different conclusion now.

In both experiments, Dror's results show that examiners can be made to disagree with themselves by extraneous information given to them when they repeat the examination years later. These results suggest that application of the method by a skilled examiner fails to determine his conclusion. Instead, information irrelevant to the comparison process biases the conclusion.

Evaluation Stage: Wertheim, Langenburg & Moenssens (2006)

In the context of week-long training courses, 92 working fingerprint examiners were given packets of prints each containing eight suspect ten-print fingerprint or palm print cards, and ten latent prints. All of the latent prints were of value, and each matched one of the suspects. The instructor adjusted the difficulty level of each packets to the examiner's performance. The examiner was asked to record an identification for each latent he could, and leave the response form blank if he could not. Only 2 erroneous identifications were reported (well less than 1%), but since the authors did not report the number of missed identifications (they did not score the blank responses), there is no way to determine a denominator for a proper error rate. As different problems, all of the latents had a matching exemplar: there were no exclusions. The examiners were not required to keep bench notes of their work, so it is impossible to know which version of ACE was used, if any. The information content of the latents was high (all were of value), difficulty was adjusted to each examiner's skill (these were training exercises, not tests), examiners could ask the instructor for hints which were freely given, and they could choose freely without penalty which latents to leave blank. For these reasons, these data do not permit a generalization of examiner accuracy rates in casework.

Evaluation Stage: Langenburg, Champod & Wertheim (2009)

Forty-three expert fingerprint examiners were given six latent-exemplar pairs to compare, half of which were ground truth one donor. The testing context was an educational workshop at a conference; those attending were probably above average in skill, and they knew they were being tested. While the latents were supposed to vary in difficulty, no basis was given for their selection, and all were judged to be of value (well above average latents). The examiners were given a worksheet and asked to conclude, for each pair, identification, exclusion, or inconclusive. The overall results showed that 218 of the 257 responses were correct, or 85%. Of the remaining 39 responses, one was an erroneous identification (0.4%), 3 were erroneous exclusions, and 35 were inconclusive. The examiners were not required to show their work, so there is no evidence of the methods they employed. The results as reported do not permit a complete picture needed for an error rate computation, and the lack of correspondence to casework latents means that whatever error rates are reported cannot be generalized to examiner accuracy in casework..

Evaluation Stage: Langenburg (2009)

Six expert examiners were given a set of sixty latent fingerprints, and a set of eight exemplar finger and palmprints. Permitted responses were identification, exclusion, inconclusive and no value. Of the 360 definitive responses, 278 were correct (77%). Of the remaining 82 responses, one was an erroneous identification (0.4%), 3 were erroneous exclusions, 41 were inconclusive, and 37 were of no value. However, the difficulty level of these latents was neither quantified nor controlled, but evidence of the time to complete each comparison (about 8 minutes) suggests quite easy latents. The examiners were not required to show their work, so the method used is not in evidence. The results are not reported in a way that permits calculation of missed identification, missed exclusion or discrimination rates. For these reasons, the results cannot be generalized to examiner accuracy rates in casework.

Evaluation Stage: Smith (2004)

The subjects were six working examiners in a single crime laboratory. Latents of varying difficulty were given, some rated as easy and some difficult, and the examiners had to complete a full analysis of the latent, followed by comparison to a number of exemplars. A correct exemplar was always present, so correct exclusion rates were not assessed. In the set of latent prints of less difficulty, 51% of the exemplars were correctly identified, 48% were missed identifications, and 1% erroneously identified. With more difficult latents, only 18% of the exemplars were correctly identified, 79% were missed identifications, and 3% were erroneously identified. Like the Evett and Williams (1996) results, Smith's results indicate a very low rate of erroneous identifications purchased at the cost of a very high rate of missed identifications.

The studies we have reported here document examiner disagreement in the Analysis, Comparison and Evaluation stages of the ACE method. The ACE method is being applied unreliably. Whenever disagreement occurs, only one answer, at most, can match ground truth. The rest are erroneous, implying invalidity. Since the examiners tested in these experiments were all highly trained and experienced, these results implicate errors inherent in the ACE method, not in the skill of the examiner. None of these studies by themselves allows a computation of an error rate for the ACE method, because the experimenter did not ascertain if the subjects in the experiment used the ACE method, or used it correctly. However, the results show that the inconsistency among examiner responses across the steps of the method suggests a substantial methodological error rate.

Opinion 2: Erroneous Identifications Testified to in Court Suggest a Substantial Error Rate.

Cole (2001, 2005), and Haber & Haber (2004, 2007, 2009, chapter 8) collected reported all of the published examples they could find of court-attested erroneous identifications through 2005, including their verifications by fingerprint examiners. The list included 61 erroneous identifications made in 22 different cases. Cole (2005) provides a number of reasons why this list underestimates by an unknown amount the true number of erroneous identifications. Morris (2009) brought the records up to 2008, adding a dozen more recent erroneous identifications that survived court, as well as analyses of a half dozen instances of individual and laboratory-wide fraud in testimony.

These data cannot be used to estimate an overall error rate for the fingerprint profession, nor for any particular method. However, they suggest that errors do occur, and that courts are not foolproof in catching errors made by examiners in the laboratory.

E. OPINIONS REGARDING THE FORENSIC FINGERPRINT PROFESSION'S LACK OF QUALITY CONTROL.

Every profession that demands its products meet standards for accuracy has to protect its work against the intrusion of errors and avoid procedures likely to create errors. As research scientists, we have found systemic evidence that the fingerprint profession fails in both respects.

There is no single regulatory organization that represents and regulates fingerprint examiners and the crime laboratories in which examiners work. Without these organizations, it is to be expected that quality controls are lacking or absent. Standards

and quality controls for the selection and training of fingerprint examiners are absent. Proficiency and certification testing are not regulated and the tests lack evidence of validity and reliability. Quality control standards for the accreditation of crime laboratories employing fingerprint examiners apply to only a small percent of laboratories, and are not enforced in those to which they do apply. The profession fails to address bias and its consequences of increased erroneous judgments. Examiners are not required to document comparison procedures and the bases of conclusions.

Opinion 1: The Fingerprint Profession is not Regulated by a Unitary Organization.

First Basis: There is no census of fingerprint examiners, and no single professional organization to which all of them belong. The IAI has a membership of approximately 7,000 forensic practitioners, of whom somewhat less than half label themselves as fingerprint examiners. Fitzpatrick (2008) an official of the IAI, acknowledged that the IAI's membership did not include all examiners working in the United States. The NAS report (2009) bemoaned that there was no accounting of the number of fingerprint examiners. The population of forensic fingerprint examiners is undocumented and therefore presently cannot be regulated.

Second Basis: Neither the IAI nor any of its subdivisions and affiliates has the authority to make regulations that are binding on fingerprint examiners or fingerprint examination procedures. All statements of regulations, rules, or procedures are "recommendations." Neither the federal government nor any national organization has responsibility or control over fingerprint examiners or examination procedures.

Third Basis: The NAS report (2009) contained a strong recommendation that a new governmental institute be created by Congress to serve this function. The IAI has not supported this recommendation (Identification News, 2009, vol. 39,3, p.9), and has proposed in its place a consortium of forensic organizations, none of whom would have any regulatory control over the fingerprint profession.

Opinion 2: There are no Standardized Regulations and Requirements Pertaining to Fingerprint Examiners.

Fingerprint examiners are entrusted to make life-threatening decisions based on substantial expertise that must be acquired, practiced, reviewed, updated and continually tested. However, the requirements for fingerprint comparison training do not resemble any of the other professions on which our society depends for its safety or health. There are no requirements. Specifically, it is our opinion that there is no standardized required educational or experiential background to be admitted into a training program, there are no standardized required training programs for fingerprint examiners, there are no standardized requirements for training (only advisory recommendations), there is no evidence that whatever training is provided constitutes a basis for professional standards of competence, there are no standardized requirements that a working examiner be certified, and there are no standardized requirements for an examiner to testify in court. We have six bases for this opinion.

First Basis: There are No Standardized Requirements for Prior Education or Experience to be Hired as a Trainee.

A few states, such as Illinois, and a few crime laboratories, such as the FBI, require specific kinds of educational background and/or experience to be hired as a trainee in latent fingerprints. However, the fingerprint profession as an entity has no formal requirements that must be met (TWGFAST, 1998), such as specific college courses, college degrees, or prior experience. Each crime laboratory can decide what qualifications it requires to hire a trainee, if any.

Second Basis: There are No Required Training Curricula, Training Manuals, or Assessments of Training Accomplishment to Become a Fingerprint Examiner.

The fingerprint profession has published a number of recommendations (not requirements) regarding training (SWGFAST, 2002c; 2006; TWGFAST, 1998), but there are no formal, standardized requirements for training to become a latent fingerprint examiner. The profession does not have a required curriculum. The profession has not created a standardized and required training manual. The profession has not developed standardized tests to use during the course of training to show that training has been mastered to a given level of competency.

At present, the majority of working examiners in crime laboratories received their training on-the-job, from more experienced examiners who themselves did not go through a standardized training program. As a result, examiner performance is expected to vary. This variation is found in all of the comparison experiments published (reviewed in Section F, and by Cole, 2005; Haber & Haber, 2004; 2009, chapter 8), on the CTS and ASCLD proficiency test results (Cole 2005), the IAI certification results (Cole, 2005), and the survey of examiners' conclusions for the Mitchell case (Haber & Haber 2004).

Third Basis: There is No Required Supervision Period Prior to Becoming a Working Fingerprint Examiner.

The decision that a trainee is competent to become a working examiner is often made informally by the more senior examiner doing the training, without a formal test. Alternatively, after a trainee passes a final examination, a crime laboratory may transition him directly into performing casework independently. No regulations require an initial period of close supervision, and supervision varies from agency to agency.

Fourth Basis: There is No Requirement for Annual Proficiency Testing of all Examiners.

Only about 600 persons are tested a year in the CTS proficiency testing program. Some of these persons may not be fingerprint examiners, some may not be employed as fingerprint examiners, and some may not be members of the IAI. Since the IAI has around 2,500 active members who describe themselves as fingerprint examiners, the annual testing program, if all of the test-takers were members, includes less than 25% of the IAI membership. The total percentage of working examiners who are tested for proficiency annually could be less than 10% (Haber and Haber, 2009, chapter 7).

Fifth Basis: There is No Requirement for All Examiners to be Certified.

The profession does not require fingerprint examiners to be certified and the vast majority of working examiners are not certified. Only 30% of the IAI members who define themselves as fingerprint examiners working in this country are certified (Haber & Haber, 2009). An unknown, large additional number of forensic fingerprint examiners are not members of the IAI, so 30% is an over-estimate of the certification rate. The National Academy of Sciences Report (2009, section S-19, Recommendation 7) states: "No person (public or private) should be allowed to practice in a forensic science discipline or testify as a forensic science professional without certification." The California Crime Laboratory Review Task Force on crime laboratories made a similar recommendation (State of California, 2010). These recommendations have no regulatory power.

Sixth Basis: There are No Qualification Requirements for a Fingerprint Examiner to Testify in Court.

The fingerprint profession has no requirements that an examiner must satisfy to offer testimony in court on a fingerprint comparison. For example, the examiner need not be proficiency tested or certified for the profession to allow the examiner to represent the profession in court.

Opinion 3: Crime Laboratories that Perform Forensic Fingerprint Work Are Not Regulated by a Single Unitary Organization.

First Basis. There is no census of laboratories employing fingerprint examiners, and no single professional organization to which all laboratories belong.

Fitzpatrick (2008) estimated that about 8,000 organizations in the United States provide forensic fingerprint comparison evidence through the employment of fingerprint examiners. The American Society of Crime Laboratory Directors (ASCLD) has about 350 members; somewhat over 300 of these laboratories are accredited. ASCLD (2006) reported that two-thirds of fingerprint identifications are made outside traditional crime laboratories. The population of forensic laboratories in which fingerprint comparisons are performed is undocumented and therefore cannot be regulated.

Second Basis. Accredited Laboratories are Not Required to Comply with ASCLD Recommendations.

The laboratory accreditation procedures, which pertain only to the few laboratories they accredit themselves are recommendations, not requirements, and they are not enforced. There is no assurance that even the small number of accredited laboratories have quality control procedures in place that reduce errors or prevent mistakes. As one example, ASCLD (1998) recommends, as an important qualification, that for a laboratory to receive and maintain accreditation by ASCLD, every latent print examiner have a BA degree. However, ASCLD does not publish information about the number of examiners working in accredited laboratories who have BA degrees, and because this is a recommendation and not a requirement for accreditation, it is not enforced.

Opinion 4: There are No Standardized Regulations and Requirements Pertaining to Crime Laboratories.

Laboratories that perform measurements need quality controls in place to insure the highest quality product possible. The quality controls assure that procedures be followed, that supervision be present, and that contamination be controlled or eliminated. In our opinion, such quality control procedures are not required or in place in most laboratories performing fingerprint comparisons. We have four bases for this opinion.

First Basis. The Profession Lacks Standard Operating Procedures (SOPs) Regulating the Work Performed in the Crime Laboratory.

SOPs for a crime laboratory would include, for example, regulations determining work flow, how a fingerprint comparison is performed, verification procedures, the requirements for bench notes and their contents, the requirements for case reports and their contents, requirements for supervision and for case reviews, and for the investigation of errors, and procedures to prevent bias. No standardized SOPs have been put in place and required by the profession, so that SOPs vary from laboratory to laboratory. Many laboratories have no quality control manuals.

Second Basis: Variation in Laboratory Procedures Means an Individual Laboratory Cannot Assure the Court of the Quality of its Product.

The absence of standardized regulations governing crime laboratories means laboratories vary widely in the procedures they follow. The absence of SOPs within an individual laboratory means that different technicians vary in the way they perform their duties. The quality of the work product will vary accordingly. An assessment of today's work product does not predict the quality of tomorrow's. The National Academy of Sciences Report (2009, section 7-19, Recommendation 8) recognized the lack of quality control regulations for crime laboratories and the need for correction: "Forensic laboratories should establish routine quality assurance and quality control procedures to ensure the accuracy of forensic analyses and the work of forensic practitioners." This requirement particularly applies to crime laboratories performing fingerprint comparisons.

Third Basis: Evidence of Fraud at the Level of the Entire Laboratory Implicates Poor Laboratory Quality Control.

Currently, more than a half dozen crime laboratories in major cities of the United States are under investigation for the systematic introduction of fraudulent or erroneous testimony in Courts (Morris, 2009). In some of these investigations, hundreds of cases are being questioned.

This epidemic can only happen in the absence of quality controls over the work in a laboratory. This is a professional failure, in which the profession and the laboratory have no control over the quality of work performed, and no protocols to assess errors after they are made.

F. OPINIONS REGARDING CONSEQUENCES OF LACK OF QUALITY CONTROLS.**Opinion 1: Fingerprint Proficiency and Certification Tests are Useless as Assessments of Examiner Skill.**

Proficiency test results have multiple uses, and the intended use of the test determines its content (what is tested, the difficulty level(s) of the items) and how it is scored. The profession itself has not stated the precise function a proficiency test should serve. Our opinion is that the current proficiency test offered through Collaborative Testing Services (CTS) serves no function at all: the test results are irrelevant to the accuracy of the ACE method; they do not permit an estimate of examiner accuracy overall, as a profession; nor do the results permit an evaluation of the skill of an individual examiner.

CTS explicitly states that the test results are not intended to represent the performance accuracy of the profession: "...The results compiled in the Summary Report are not intended to be an overview of the quality of work performed in the profession and cannot be interpreted as such..." (CTS, 01-516). This proviso does not boost confidence in the testing program.

Proficiency testing of fingerprint examiners has been in limited use since the 1970s. In the United States, ASCLD contracted with CTS to manufacture and score latent fingerprint comparison tests. Presently, about 600 persons take these tests annually. The test assesses only latent-to exemplar comparisons. Each examiner receives 10-12 latent prints and several suspect exemplar prints or full tenprint cards, and the examiner is asked to identify some or all of the latents to one of the exemplars, or mark them as not identified. In some years, every latent has a matching exemplar; other years one or two latents do not.

First Basis: These Test Results Cannot be Generalized to Normal Casework.

With only a few exceptions over the 30 years, each latent to be compared is of value (much easier than casework, where an estimated 75% of latents are of no value for comparison); no comparisons based on an AFIS search have been included on these tests (about half of casework does not have an initial suspect and requires a computer search); the matching suspect's exemplar usually or always is provided (producing an expectation that the suspect's exemplar is present, whereas nearly all casework comparisons involve a different expectation, because most conclusions are exclusions). This mismatch between normal casework and the test used to predict the quality of that work means that the results of the test cannot be generalized to accuracy in normal casework.

Second Basis: The Difficulty of the Latent Prints Used as Test Items is Unknown.

Latent prints vary in quantity and quality of information content (Ashbaugh, 1999). No research has ever been reported that measures the information content (difficulty) of a latent. While the fingerprint profession recognizes that this objective scale can be constructed (FBI, 2007), it has never been developed (also Section C, Opinion 1). Until an objective scale of information content from difficult to easy latent fingerprints has been validated and shown to be reliable, proficiency test results cannot be interpreted. Results from different tests cannot be compared, and results from different years cannot be compared. It is also

impossible to show that the test item latents resemble casework latents. Without this assurance, test results cannot be generalized to casework accuracy.

Third Basis: The Fingerprint Profession Has Not Defined a Level of Skill Needed to Pass the CTS Proficiency Test.

The CTS proficiency test results are reported as raw scores. The profession has not established what constitutes a passing score. The skill level needed to be “competent” or proficient is undefined.

Fourth Basis: The CTS Proficiency Test Does Not Differentiate Between Good and Poor Examiners.

The overall CTS proficiency test scores over the past 7 years have ranged between 87% and 97% correct, with an average well above 90%. When virtually every examiner gets close to a perfect score, the test results cannot discriminate good from poor examiners. Because there is so little variation in scores, the test results are not informative about examiner proficiency.

Fifth Basis: The CTS Proficiency Test Does Not Include Assessment of the Method or Features Used to Reach Conclusions.

Bench notes have never been required for the CTS proficiency test. Therefore, the test assesses only conclusions, but not examiner accuracy in reaching those conclusions. The CTS test does not assess ACE.

Sixth Basis: The Proficiency Tests Use Incorrect Response Categories, Greatly Reducing the Power of the Test Results.

The rules of the fingerprint profession (SWGFAST, 2004) restrict examiners to four conclusions: the latent is of *value* or *no value* for comparison; the suspect can be *excluded* as the donor of the latent; the suspect can be *individuated* as the donor of the latent; or the comparison is *inconclusive*. On the CTS proficiency tests, the no-value response is eliminated and the insufficient category is not allowed. On these tests, only two conclusions are allowed, identified and not identified. Not identified is ambiguous: could the examiner exclude the latent, or was he unable to decide? The non-conventional response categories prevent generalization of the results to casework. Only one of the four conclusions an examiner routinely makes in casework can be assessed for accuracy (identification). Examiner accuracy in reaching other conclusions cannot be scored.

Seventh Basis: No Assessment of the Validity or Reliability of the Proficiency Tests or the Certification Tests Have Ever Been Reported.

Validity: do proficiency tests actually measure examiner skill? If so, the scores on the proficiency test would correlate with other reasonable measures of proficiency, such as the certification test scores, scores on training tests, or supervisor ratings. No evidence of this kind, which could serve to establish the validity of the tests, has ever been published. Given the deficiencies in the test noted above, it is likely that the proficiency tests do not assess skill very accurately, and therefore have very poor validity.

Reliability: when a test's reliability is high, it indicates three important properties. (1) The score achieved by an individual examiner can be trusted: he's likely to get the same score if he takes a comparable test again. (2) Differences in scores between individuals can be trusted to be real differences and not just chance due to measurement errors in the test. (3) Differences in scores as a function of different amounts of training or experience can also be believed. None of these is true if the test isn't reliable. Reliability results have never been reported for the CTS proficiency test.

Eighth Basis: Scoring and Sampling Procedures Underestimate the Error Rate.

When an examiner takes a proficiency test alone and unaided, and records his answers on his own answer sheet, the score he receives is his alone. However, if two or more examiners work together (as is a frequent practice in a crime laboratory), the error rate recorded for their combined responses is much lower than their error rate if they had worked independently. For example, assume that two examiners, each working individually, has a 14% erroneous identification rate. When they work together, they report an identification only when they both agree it is an identification. Their individual 14% error rate is reduced to only 2% because they catch some of each other's errors (see Haber & Haber, 2009, chapter 8 for more examples). Therefore, the scores of examiners working together underestimates by a substantial amount the proficiency scores of the same individual examiners working independently. The current scoring of these tests combines scores from examiners who worked independently with those who worked in groups, so that the scores reported are inflated. As the example above shows, the inflation rate is substantial.

The profession provides no information about the sample of examiners who take the test. The number of examiners who participate in CTS and ASCLD testing each year is a tiny proportion of all working examiners (Haber & Haber, 2009, chapter 7), but if only good examiners take advantage of proficiency testing, their average scores will overestimate the performance of the entire population of examiners. As one example, ASCLAD recommends annual proficiency testing for a laboratory to maintain accreditation. While only a few laboratories are accredited, those are likely to be better laboratories with better trained examiners. It is those examiners who are more likely to undergo proficiency testing, so the sampling bias is likely to be severe.

Ninth Basis: Judicial Notice has been Given of Inadequate Proficiency Testing.

In the only review of fingerprint proficiency tests by a Federal Court (*USA v. Plaza II*, 2002), the court conducted a Daubert hearing on the adequacy of the CTS proficiency tests sponsored by the IAI and the ASCLD, and the in-house FBI proficiency tests. Based on the evidence presented, Judge Pollak dismissed these tests as unacceptable to a court on the grounds that these tests cannot claim to measure the quality of performance of latent print examiners, including their error rates. Cole (2005) has provided a detailed interpretation of Judge Pollak's assessment.

Opinion 2: Examiner Exposure to Bias in Crime Laboratories Increases the Probability of Erroneous Judgments.

In our opinion, the fingerprint profession has failed to require the laboratory to maintain quality controls to prevent bias: controls that are an assumed standard in other governmental assessment programs such as drugs, environmental hazards, or transportation safety. The purpose of bias control is to prevent a decrease in examiner accuracy of perception, measurement and judgment. We have eight bases for this opinion.

First Basis: Bias Reduces Accuracy.

Ashbaugh (1999), Dror et al., (2008), Risinger (2002) and the Office of the Inspector General Report (Fine, 2006) emphasize a number of ways in which irreversible and detrimental effects of bias and contamination can intrude into the application of ACE. Bias has been shown to increase erroneous conclusions.

The contribution of bias to erroneous conclusions is documented in a widespread research literature spanning over a century (Risinger, et al. 2002), is explicitly acknowledged as having contributed to the FBI erroneous identification of Brandon Mayfield's fingerprints (Budowle, et al., 2006), and has been shown in research results on fingerprints (Dror & Charlton, 2006; Dror et al., 2006; Dror & Frasier-macKensie, 2008). Knowledge about the seriousness of the crime, previous history of the suspect, or the presence of other evidence in the case biases the examiner's comparison process, as does knowledge that another examiner individuated the comparison (non-blind verification).

Second Basis: Quality Controls to Prevent Bias are Absent in Crime Laboratories.

ASCLD, SWGFAST, TWGFAST and IAI recommendations contain no procedures to prevent bias. The National Academy of Sciences Report (2009, S-18, Recommendation 5) recognized this lack and saw correction as critical: "... [The] National Institute for Forensic Sciences should develop standard operating procedures (that will lay the foundation for model protocols) to minimize, to the greatest extent reasonably possible, potential bias and sources of human error in forensic practice."

Third Basis: Prior to Examination, Examiners are Exposed to Biasing Information.

In crime laboratories that serve as part of a police department, the examiner is routinely aware of contaminating information extraneous to the fingerprint comparison task. This information includes whether the suspect has a prior record, the nature and seriousness of the crime (Dror et al, 2005; Stacy, 2004; Fine, 2006), other evidence against the suspect (Dror, Charlton & Peron, 2006), and how certain the investigators are that the suspect is the perpetrator. Exposure to this biasing information occurs before the examiner sees the latent print and begins analysis, and increases the probability that the examiner will make an erroneous identification.

Fourth Basis: During Analysis, Examiners are Exposed to Bias.

Examiners are permitted to view exemplars before completing Analysis. Exemplars are usually larger than the latent, less distorted and have mostly unambiguous features. When an examiner views an exemplar, he usually can determine its features easily and accurately. When he then looks at an ambiguous latent print that has some resemblance to the exemplar, it is relatively easy to "see" in the distortions in the latent some of the features clearly displayed in the exemplar.

Dror, Charlton and Peron (2006) and the Office of the Inspector General (Fine, 2006) document the changes viewers make in the details they "see" in a latent on the basis of knowledge of the exemplar. Using the exemplar to define the latent biases the results towards an erroneous identification, because the exemplar's friction ridge pattern is being used to resolve ambiguities in the latent.

Ashbaugh (1999) argues that any exposure to an exemplar before the analysis is done makes it easier to find agreements when they would not have been noted or accepted if the examiner had started with the latent first. Whenever an examiner has perused the exemplar prior to fully completing analysis of the latent, to the extent that confirmation bias intruded, the validity of conclusions based on the method is reduced.

Fifth Basis: During Comparison Examiners are Exposed to Bias.

In Ashbaugh's (1999) explication of the ACE method, the examiner is required to complete an entire analysis of the latent print before proceeding further. Ashbaugh describes different circumstances where bias can intrude if the analysis is left unfinished, most of which result from uncertainty about distortions after comparison begins. The only way to prevent them, he argues, is to complete the analysis, including careful assessment of all uncertainties or ambiguities, so that later examination of the exemplar during comparison does not allow the examiner to change what he found during the analysis. Otherwise, backward reasoning (Ashbaugh, 1999) and confirmation bias (Risinger, et al., 2002) can lead to incorrect conclusions during the comparison and evaluation stages of ACE-V. In the absence of laboratory requirements for a complete analysis of the latent as the first step, with accompanying documentation of compliance, conclusions based on ACE have a lower validity than if the Analysis stage had been properly completed first.

Sixth Basis: During Evaluation, Examiners are Exposed to Bias.

The most serious contaminating information that biases the outcome of evaluation is examiner expectation that the exemplar matches the latent. This bias leads the examiner to overlook discrepancies and accept less agreement than he normally demands.

Dror's research and the Office of the Inspector General (Fine, 2006) report examples. Until laboratory controls to prevent examiner exposure to biasing information are instantiated and required, the validity of conclusions based on ACE-V is reduced.

Seventh Basis: During Verification Examiners are Exposed to Bias.

Blind verification is currently neither required nor practiced. In the absence of blind verification, the validity of conclusions based on ACE-V is reduced by bias. Non-blind verification introduces bias. Erroneous conclusions are less likely to be discovered when the verifier knows what conclusion had been already reached (Arvizu, 2002; US Food and Drug Administration; Boone, et. al., 1982). The FBI (Budowle, et al., 2006; Srmz, et al., 2006), and the Office of the Inspector General (Fine, 2006) have recently argued that blind verification should be required to avoid bias and improve the validity of conclusions. The NAS report also strongly noted that non-blind verification was not a useful way to catch errors.

Eighth Basis. The Lack of Blind Verification has Received Judicial Notice.

In *New Hampshire v. Langill*, 2007, the trial court judge excluded the fingerprint evidence because the conclusion was not verified blindly. While this ruling was subsequently overturned (2008), the basis for blind verification was upheld.

Opinion 3: The Use of AFIS Contains Many Bias Traps.

AFIS is used in about half of all cases in which fingerprint evidence is collected from the crime scene (Haber & Haber, 2009, chapter 6). Because its use puts the examiner at risk of exposure to bias, bias protections are needed. They are currently absent. We consider three sources of bias.

First Basis: Bias from Seeing Exemplars Before Completing Analysis of the Latent.

Dror, Charlton & Peron, (2006) and the Office of the Inspector General Report (Fine ,2006) both document the changes viewers make in the details they “see” in a latent on the basis of knowledge of the exemplar. AFIS typically does not require complete analysis of the latent for submission. Consequently, the examiner receives the AFIS exemplars without having completed an analysis of the latent print. Ashbaugh (1999), and the Office of the Inspector General Report (Fine, 2006) both strongly emphasize the bias introduced by examination of the exemplar prior to a complete analysis of the latent. Using the exemplar to define the latent biases the results towards an erroneous identification. The Office of the Inspector General Report (Fine, 2006) documents the danger of this procedure in its analysis of the FBI erroneous identification of Brandon Mayfield.

When a set of suspect exemplars is produced by AFIS, most examiners report that they quickly skim the set of candidates to exclude those candidates with a different level 1 pattern or other gross differences. If one or more candidates appear likely, the examiner checks those further against the latent. As a result, he knows a number of the features of an exemplar before he has completed analysis of the latent. This practice biases the examiner with information acquired from an exemplar.

Second Basis: Bias from Knowing Ranking/Rating Information.

AFIS outputs include a ranking and/or rating of the amount of similarity between the processed latent and each candidate. Rankings are both unreliable and changeable. They vary as a function of the search-similarity algorithms of the AFIS, the size of the data base, and the preprocessing of the latent. Ranking/rating information biases the examiner in three ways. He is more likely to compare higher ranked candidates first, even in the order of ranking; he is likely to devote more effort to them; and because examiners believe their identifications are “gold standard,” once an examiner makes an identification, he stops searching. Each of these increases the chances of an erroneous identification.

Third Basis: Bias from Reconfiguring an Input to AFIS.

The majority of AFIS searches do not produce a hit. Some diligent examiners try again. To do so, they alter the details they had submitted to the search system the first time. A second strategy is to submit the latent to a different database. The National Academy of Sciences Report (2009, section 10-4) asserts: "In fact, experienced latent print examiners have found that different systems will retrieve different prints in response to a given input map of features, and they have learned system-specific ways of annotating features on a latent print in order to maximize the success of each system's (inferred) search algorithms." In other words, the examiner adjusts his interpretation to the system.

An AFIS failure to produce a match arises for three reasons. The most likely is that the perpetrator is not in the database. A second possibility is that the AFIS algorithms are not powerful enough to find the true exemplar in the database. For either reason, the AFIS output is the most useful the examiner can get, and further attempts can only produce error.

A third possibility is that the examiner mislabeled characteristics of the latent when he encoded it for submission. If this happened, isn't this diligent examiner more likely to get the features right the second time? The answer is No. When he altered the details of the latent print to repeat the search, he increased the probability of finding a set of erroneous candidates, because he used less likely features to input. If the examiner alters the input enough times, eventually he'll get a candidate with sufficiently similar characteristics for an identification. The probability of a correct match decreases with each reconfiguration of the latent.

IV. SUMMARY OPINION

As research scientists, who also are trained in fingerprint comparisons, it is our opinion that the validity of the Analysis-Comparison-Evaluation method of fingerprint comparison has never been tested scientifically and its error rate is unknown. The method is untestable until the profession agrees in a description of the features in a fingerprint to use for comparison, how to specify their location objectively, how to quantify the information content in a fingerprint, how to describe distortion objectively, and how to measure similarity objectively.

If the validity of the method is to be testable, the profession must set objective standards for latent print difficulty (value), distortion versus discrepancy (exclusion) and sufficient similarity for identification. These standards at present are subjective and differ from examiner to examiner.

Fingerprint examiners believe that their work meets scientific standards, and that when they make an identification they are right. There is no more evidence today to support these beliefs than there was 100 years ago.

Nonscientific procedures or beliefs cannot substitute for an empirical scientific test of the accuracy of the method. This empirical study has never been done.

V. CITATIONS OF REFERENCES

Anthonioz, A., Egli, N., Champod, C., Neumann, C., Puch-Solis, r., & Bromage-Griffiths, A. (2008). Three level detail and their role in fingerprint identifications. *J. Forensic Identification*, 58, 562-589.

Arvizu, J. (2002). Testimony on Mr. Plaza's motion to exclude the government's latent fingerprint evidence, hearing before Judge Pollak, USA v. Plaza, February 24, 2002.

ASCLD 2006 strategic plan for training. <http://www.ascl-d-lab.org>

ASCLD (1998). Laboratory accreditation Programs. <http://www.ascl-d-lab.org>.

- Ashbaugh, D.R. (1999) *Quantitative-Qualitative Friction Ridge Analysis: An introduction to basic and advanced ridgeology*. Boca Ratan, FL: CRC Press.
- Ashbaugh, D.R. (2005). Proposal for ridges-in-sequence.
- Babler, W.J. (1991). "Embryologic development of epidermal ridges and their configuration." *Birth Defects, Original Article Series*, 27, 95-112.
- Beeton, M. (2002). "Friction ridge identification process--Purposed scientific methodology." *The Detail*, 2-18-2002; <http://clpex.com>.
- Black, J. F. (2006). "Pilot Study: The application of ACE-V to simultaneous (cluster) impressions." *J. Forensic Identification*, 56, 933-
- Boone, D.J., Hansen, H.J., Hearn, T.L., Lewis, D.S., & Dudley, D. (1982). Laboratory evaluation and assistance efforts: mailed, on-site and blind verification proficiency testing: surveys evaluated by the Center for Disease Control. *Am. J. Public Health*, 72, 1364-1368.
- Budowle, B., Buscaglia, J., & Perlman, R.S. (2006). "Review of the scientific basis for fingerprint comparisons as a means of identification. Committee Findings and recommendations." *Forensic Science Communication*, 8, 1-16.
- California Crime Laboratory Review Task Force (2010). Final report: An examination of forensic science in California. State of California Publications.
- Champod, C. (2008). Fingerprint examination: Toward more transparency. *Law, Probability and Risk*, 7, 111-118.
- Champod, C., Lennard, C. Margot, P. & Stoilovic, M. (2004.) *Fingerprints and other friction ridge skill impressions*. Boca Ratan, FL: CRC Press.
- Collaborative Testing Services (CTS). www.collaborativetesting.com.
- Cole, S. A. (2001). *Suspect Identities: A history of fingerprinting and criminal identification*. Cambridge, MA.: Harvard University Press.
- Cole, S.A. (2004). "Grandfathering evidence: Fingerprint admissibility rulings from Jennings to Llera Plaza and back again." *American Criminal Law Review*, 41, 1189-1276.
- Cole, S.A. (2005). "More than zero: Accounting for error in latent fingerprint identifications." *J. Criminal Law and Criminology*, 95, 985-1078.
- Cole, S.A. (2008). "Comment on Haber & Haber (2008)". *Law, Probability Risk*, vol. 7.
- Cole, S.A., Welling, M., Dioso-Villa, R., & Carpenter, R. (2008). "Beyond the individuality of Fingerprints: a measure of simulated computer latent print source attribution accuracy." *Law, Probability and Risk*, 7, 165-189.
- Cowger, J.F. (1993). *Friction ridge skin: comparison and identification of fingerprints*. Baco Ratan, FL: CRC Press
- Dror, I.E., & Charlton, D. (2006). "Why experts make mistakes." *J. Forensic Identification*, 56, 600-616.

Dror, E., Charlton, D., & Peron, A. (2006). "Contextual information renders experts vulnerable to making erroneous identifications." *J. Forensic Identification*, 56, 74-78.

Dror, I.E., & Fraser-MacKensie, P.A.F. (2008). "Cognitive biases in human perception, judgment and decision making: bridging theory and the real world." In K. Rossmo (Ed.), *Criminal Investigation Failures* (Ch 5, pp 53-67). London: Taylor & Francis.

Dror, I.E., Peron, A., Hind, E., & Charlton, D. (2005). When emotions get the better of us: The effect of cognitive top-down processing in matching fingerprints. *Applied Cognitive Psychology*, 17, 799-808.

Evetts, Z.W., & Williams, R.L. (1996). "Review of the 16 point fingerprint standard in England and Wales." *J. Forensic Identification*, 46, 49-73.

Federal Bureau of Identification (1988, last edition). *The Science of Fingerprints: Classification and uses*. United States Department of Justice

Federal Bureau of Identification. (2007). A research program on latent print difficulty and dimensions. Announced at the 2007 IAI Education Conference, San Diego, CA.

Fine, G.A. (2006). *Review of the FBI's handling of the Brandon Mayfield case*. Washington: US Department of Justice, Office of the Inspector General.

Fitzpatrick, F. (2008). "Whither Accreditation." *Identification News*, 38, 9-10

Food and Drug Administration (2001). Guidance for Industry: Choice of control group and related interests in clinical trials. Washington: FDA (29 pages).

Galton, F. (1896). *Finger Prints*. London: McMillin.

Garrett, R.J. (2009) Editorial Comments. *Identification News*, vol 39;

Haber, R.N. (2002). Testimony on Mr. Plaza's motion to exclude the government's latent fingerprint evidence, hearing before Judge Pollak, USA v. Plaza, February 24, 2002

Haber, L., & Haber, R.N. (2004). "Error rates for human latent fingerprint examiners." In N. Ratha & R. Bolle (Eds.), *Automatic Fingerprint Recognition*, pps 339-360. New York: Springer Verlag.

Haber, L. & Haber, R.N. (2008). "Scientific validation of fingerprint evidence under Daubert." *Law, Probability and Risk*, 17, 87-

Haber, L., & Haber, R.N. (2009). *Challenges to Fingerprints*. Tucson: Lawyers and Judges Publishing Co.

Hall, L.J., & Player, L. (2008). With the introduction of an emotional context affect fingerprint analysis and decision-making. *Forensic Science International.*, 181, 36-39.

International Association for Identification (1973) "Standards committee report." *Law Enforcement Bulletin*, pp 7-8.

Interpol European Expert Group on Fingerprint Identification (IEEGGI II) (2005). "Method for fingerprint identification." [http:// www.interpol.int/public/forensic/fps/working_parties](http://www.interpol.int/public/forensic/fps/working_parties).

Langenburg, G. (2004). "Pilot study: a statistical analysis of the ACE-V method-analysis stage." *J. Forensic Identification*, 54, 64-74.

Langenburg, G. (2009). "A performance study of the ACE-V process: A pilot study to measure the accuracy, precision, reproducibility, repeatability and bias ability of conclusions resulting from the ACE-V process." *J. Forensic Identification*, 59, 219-257.

Langenburg, G., Champod, C., & Wertheim, P. (2009). "Testing for potential contextual bias effects during the verification stage of ACE-V methodology when conducting fingerprint comparisons." *J. Forensic Science*, 50, 1-12.

Leo, W. (1998). "Distortion versus dissimilarity in friction ridge skin." *J. Forensic Identification*, 48, 125-129.

Maryland v. Bryan Rose (2007) Case # K-06-0545, Baltimore County Circuit Court, Oct, 19, 2007.

Martin, K.F. (2007). The International Association for Identification's position and recommendations to the National Academies of Sciences Committee to Review Forensic Science. *Identification News*, 38, 6-11.

Meagher, S. B., Budowle, B., & Siesig, D. (1998). "50K v. 50K fingerprint comparison test." In Daubert hearing before Justice Joiner, July 8-9, 1999 US v. Bryon Mitchell, US District Court PA (3d).

Meagher, S.B. (1998). Report of the Federal Bureau of Identification's Mitchell survey. In USA v. Mitchell, 365F (3d. Circuit). 1998.

Meagher, S. B. (2002). Testimony on Mr. Plaza's motion to exclude the government's latent fingerprint evidence, hearing before Judge Pollak, USA v. Plaza, February 24, 2002.

Meagher, S.B. (2003). Comments on Columbia Broadcasting System: 60 Minutes, January 5, 2003.

Mnookin, J.L. (2008). "The validity of latent fingerprint identifications: confession of a fingerprinting moderate." *Law, Probability and Risk*, 7, 127.

Morris, E. (2009 Unpublished). Update on Erroneous Convictions based on Erroneous Fingerprint Identifications. Contact information: Public Defender's Office, LA Superior Court, Los Angeles, CA.

National Academy of Sciences. (2009). *Strengthening Forensic Sciences in the United States: a Path Forward*. Washington DC: National Academies Press.

New Hampshire v. Langill (2007) (No. 050S-1129). [Http:// clpex.com/information/NH v. Langill](http://clpex.com/information/NH%20v.%20Langill)

New Hampshire v. Langill (2008). [http://clpex.com/information/NH v. langill](http://clpex.com/information/NH%20v.%20Langill).

Olsen, R.D. (1978). *Scott's Fingerprint Mechanics*. Springfield, IL: Charles Thomas.

Onin Website (2009): Legal Challenges to Fingerprints. [http:// www.onin.com/jp/daubert tests.html](http://www.onin.com/jp/daubert%20tests.html).

Risinger, D.M., Saks, M.J., Thompson, W.C., & Rosenthal, R. (2002). "The Daubert/Kumho implications of observer effects in Forensic Science: Problems of expectation and suggestion." *California Law Review*, 90, 1-56.

Smith, K. (2002). Testimony on Mr. Plaza's motion to exclude the governments latent fingerprint evidence, hearing before Judge Pollak, USA v, Plaza, Feb. 24, 2002.

Smith, R. (2004). Latent print unit evaluation. Report to Capt. Thomas Dowd, Supervision of Identification Bureau, Boston Police Dept, Sept 5, 2004.

Smrz, M.A., Gurmestein, S.G., Einseln, A., Fisher, C.L., Fran, R., Stacey, R.B., Thesin, C.E., & Budowle, B. (2006). "Review of FBI latent print unit processes and recommendations to improve practices and quality". *J. Forensic Identification*, 56, 402-434.

Souder, S. (2007). Memorandum: Ruling on Frye Hearing to Exclude Fingerprint Evidence. October, 2007. Baltimore: Baltimore County District Court.

Souder, S. (2008) Memorandum Decision Re Consideration: February 21, 2008. Baltimore: Baltimore County District Court.

Stacey, R.B. (2004). "Report on the erroneous fingerprint individualization in the Madrid training bombing case." *J. Forensic Identification*, 54, 706-718.

Stoney, D.A. (2001). "Measurement of fingerprint individuality." In H.C. Lee & Gaensslen (eds.), *Advances in Fingerprint technology* (2nd ed.), pp. 327-387. Baco Ratan, FL: CRC Press.

Swets, J.A., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1-26.

SWGFAST (2002). Friction Ridge skin examination methodology for latent print examiners. 8/22/2, ver 1.01.

SWGFAST (2002a). Fingerprint examination methodology for latent fingerprint examination. http://swgfast.org/friction_ridge_examination_methodology_for_latent_print_examination_1.01.pdf

SWGFAST (2002b). Training to competence for latent fingerprint examiners, ver 2.1, 8-22-02.

SWGFAST (2004). "Standards for conclusions." 9/11/03, v.1.0. *J. Forensic Identification*, 358-359.

SWGFAST (2006). Quality assurance guidelines for latent print examiners. 9.28.06.ver. 3.0, *Journal for Forensic Identification*, 56, 117-128.

Tangen, J.R. & Vokey, J.M. (2009). Fingerprint identification or the psychophysics of identification. *Quarterly Journal of Experimental Psychology*.

Thompson, W.C., Taroni, F., & Aitkin, C.G.G. (2003). "How the probability of a false positive affects the value of DNA statistics." *J. Forensic Science*, 46, 1-8.

Thornton, J.I. (1977). "The one dissimilarity doctrine in fingerprint identification. *J. Forensic Identification*, 56, 345-356.

Triplett, M. & Cooney, L. (2006). "The etiology of ACE-V and its present use: An exploration of the relationship between ACE-V and the scientific method of hypothesis testing." *Journal of Forensic Investigation*, 56, 245-253.

TWGFAST (1998). Minimum Qualifications for Latent Print Examiner Trainees. *Journal of Forensic Identification*, 48.

United States v. Parks (1991). Fed. Dist Court, Central Division, CA, CR-91-398 JSL.

[United States v. Plaza, 179 F. Supp. 2d 492 \(E.D. Pa. 2002\)](#) (“Plaza 1”).

[United States v. Plaza, 188 F. Supp. 2d \(E.D. Pa. 2002\)](#) (“Plaza 2”).

[United States v. Mitchell, 365 F. 3d 215, 246 \(3d Cir. 2004\)](#).

[United States v. Havvard, 117 F. Supp. 2d, 848 \(S.D. Ind. 2000\)](#).

Vokey, J.R., Tange, J.M., & Cole, S. (2008). On the preliminary psychophysics of identification. *Quarterly Journal of Experimental Psychology*.

Wells, G.L., Memon, A., & Penrod, S.D (2007). “Eyewitness evidence: improving its probative value.” *Psychological Science in the Public Interest*, 7, No.2, 45-75.

Wertheim, K, Langenburg, G., & Moenssens, A.A. (2006). “A report of latent print examiner accuracy during training exercises.” *Journal Forensic Identification*, 56, 55-92.

APPENDICES TO THIS REPORT

A. Resume for Lyn Haber, Ph.D.

B. Resume for Ralph Norman Haber, Ph.D.

C. Fingerprint publications and presentations

D. Fingerprint cases on which we have been retained

E. Copy of Haber, L., & Haber, R.N. (2007). “Scientific validation of fingerprint evidence under Daubert.” *Law, Probability and Risk*, 17, 87-109.

Appendix not available.

704 F.3d 480
United States Court of Appeals,
Seventh Circuit.

UNITED STATES of America, Plaintiff–Appellee,
v.
Clacy Watson HERRERA, Defendant–Appellant.

No. 11–2894. | Argued Sept.
28, 2012. | Decided Jan. 9, 2013.

Synopsis

Background: Defendant was charged in the United States District Court for the Northern District of Illinois, Rebecca R. Pallmayer, J., with involvement in international drug trafficking conspiracy. The same court, [James F. Holderman](#), Chief Judge, granted defendant's motion to exclude fingerprint evidence. Government appealed. The Court of Appeals, [366 Fed.Appx. 674](#), reversed and remanded. After remand, defendant was convicted and sentenced to 340 months' imprisonment. Defendant appealed.

Holdings: The Court of Appeals, [Posner](#), Circuit Judge, held that:

[1] district court's order excluding fingerprint evidence and related testimony warranted correction by mandamus, and

[2] newly assigned district court judge did not impermissibly pressure jury to complete deliberations in one day.

Affirmed.

West Headnotes (4)

[1] **Criminal Law**
🔑 [Objects used for identification](#)

Mandamus

🔑 [Criminal prosecutions](#)

In criminal prosecution for crimes arising from defendant's alleged involvement in international drug trafficking conspiracy, district court's order excluding fingerprint evidence and related testimony on grounds of government tampering

seriously disrupted prosecution's case on basis of utterly baseless but damaging imputations of grave prosecutorial misconduct, involved flouting of governing precedents, and would probably have resulted in groundless acquittal, and thus warranted correction by mandamus; evidence was excluded due to determination that government had not demonstrated requisite “chain of custody,” but government had offered 10 witnesses to establish that “chain of custody” had remained intact. [U.S.C.A. Const.Amend. 6](#).

[Cases that cite this headnote](#)

[2] **Criminal Law**
🔑 [Condition; change; tampering](#)

Once the government presents evidence that adequate precautions had been taken to preserve the evidence challenged by the defendant, it has established admissibility, though at trial the defendant can challenge the adequacy of the precautions and present evidence of tampering.

[Cases that cite this headnote](#)

[3] **Criminal Law**
🔑 [Experiments and Tests; Scientific and Survey Evidence](#)

Expert evidence, under [Daubert](#) and Federal Rules of Evidence, is not limited to scientific evidence, however such evidence might be defined; it includes any evidence created or validated by expert methods and presented by an expert witness that is shown to be reliable. [Fed.Rules Evid.Rules 702, 703, 28 U.S.C.A.](#)

[8 Cases that cite this headnote](#)

[4] **Criminal Law**
🔑 [Urging or Coercing Agreement](#)

Newly assigned judge on remand, in criminal prosecution for crimes arising from defendant's alleged involvement in international drug trafficking conspiracy, did not impermissibly pressure jury to complete its deliberations in one day, absent any evidence to support such accusation of judge noted for her patience; after instructing jury, and only moments before jury

left courtroom to deliberate, judge indicated she would make every effort to accommodate jurors' schedule, and jurors did not ask any question to judge that would have flagged concern that trial could end without verdict unless they rushed their deliberations. [U.S.C.A. Const.Amends. 5, 6](#).

[Cases that cite this headnote](#)

Attorneys and Law Firms

*481 [David E. Bindi](#), [Arianna Kastanek](#) (argued), Attorneys, Office of the United States Attorney, Chicago, IL, for Plaintiff–Appellee.

[Gabriel Fuentes](#) (argued), Attorney, Jenner & Block LLP, Chicago, IL, for Defendant–Appellant.

Before [POSNER](#), [ROVNER](#), and [SYKES](#), Circuit Judges.

Opinion

[POSNER](#), Circuit Judge.

Two years ago, in response to a petition for a writ of mandamus filed by the government during the criminal trial of the defendant on drug charges, we ordered the district court to admit into evidence an exhibit labeled “Roberson Seizure 2”; to allow the government to recall Stephen Koop to testify at trial about the recovery of latent fingerprints from that exhibit; and to allow testimony regarding comparison of the latent prints with patent fingerprints known to be the defendant's. [In re United States](#), 614 F.3d 661 (7th Cir.2010). The judge had excluded the exhibit and related testimony because he suspected, though on the most tenuous of grounds, that the government had tampered with the fingerprint evidence. He threatened to grant the defendant's request for a mistrial on the ground of prosecutorial misconduct that was (the judge believed) intended to avert a likely acquittal, a ground that if sustained would have barred any further prosecution of the defendant as *482 placing him in double jeopardy. [Oregon v. Kennedy](#), 456 U.S. 667, 679, 102 S.Ct. 2083, 72 L.Ed.2d 416 (1982); [United States v. Catton](#), 130 F.3d 805, 807–08 (7th Cir.1997); see also [United States v. Buljubasic](#), 808 F.2d 1260, 1265 (7th Cir.1987). We also ordered the case reassigned to another district judge. This was done and the trial, which had been interrupted by the mandamus proceeding, resumed, and ended shortly in the conviction of

the defendant. The judge sentenced him to 340 months in prison for a variety of drug-related offenses. He appeals.

[1] Many of his arguments repeat ones he made in the mandamus proceeding. (In effect he is asking us to rehear our previous decision—two years after the deadline for asking for rehearing expired.) The only such argument that we didn't discuss is based on [Will v. United States](#), 389 U.S. 90, 96–97, 88 S.Ct. 269, 19 L.Ed.2d 305 (1967), which forbids the use of mandamus as a substitute for an appeal that is forbidden—and the government is not permitted to appeal an evidentiary ruling in a criminal case once the trial has begun. 18 U.S.C. § 3731. But the Court in [Will](#) held only that the court of appeals hadn't explained why the district court's ordering the government to give the defendant a bill of particulars was so “seriously disruptive of the efficient administration of criminal justice in the Northern District of Illinois” as to warrant mandamus. 389 U.S. at 104, 88 S.Ct. 269. The district judge's order in the present case was no run-of-the-mill mistaken procedural or evidentiary ruling. The order seriously disrupted the prosecution's case, and did so, as we are about to show, on the basis of utterly baseless but damaging imputations of grave (criminal, really) prosecutorial misconduct; involved the flouting of governing precedents; and would probably have resulted in a groundless acquittal. The order thus warranted correction by mandamus. See [United States v. Vinyard](#), 539 F.3d 589, 591–92 (7th Cir.2008).

The chain of events that culminated in the mandamus proceeding had begun with the district judge's decision to exclude evidence that two of the defendant's fingerprints had been recovered from a bag of heroin wrapped in tape and further encased in condoms and found in a drug courier's rectum. The heroin had been removed from the bag and placed in an evidence bag and then both it and the packaging (the tape and condoms) had been placed in another evidence bag and it was this second exhibit that was at issue. The district judge's ground for excluding it was his belief that the government hadn't adequately demonstrated the requisite “chain of custody”—hadn't demonstrated that there had been no opportunity to tamper with or otherwise mishandle the evidence between the time it was obtained and the trial. The judge made this ruling in the face of the government's having offered ten witnesses to establish that the chain of custody had remained intact.

The judge was disturbed because the exhibit had, according to an evidence log sheet, gained 20 grams in weight between

May and September 2001. (Yet he attached no significance to its having gained 190 grams between September 2001 and the trial.) He thought the weight gain might have been attributable to federal officers' pressing a piece of adhesive tape containing the defendant's fingerprints (obtained elsewhere) onto the packaging of the heroin. That suspicion grew into a conviction, for which there was no rational basis, that government lawyers had lied about the chain of custody. To no avail the government explained that the reason for the increase in weight was that the bag with the fingerprints, after being opened so that the presence and amount of the *483 illegal drug contained in it could be verified, and later closed up again, had been weighed together with other bags. The reported weight was the weight of the package containing all the bags, and thus there were more bags in it. Obviously the package would not have gained 210 grams (20 + 190)—almost half a pound—from replacing a piece of the tape in which one of the bags was wrapped by a piece of tape containing the defendant's fingerprints.

The judge acknowledged that his supposition of tampering was “speculative.” That was an understatement. For among other things the defendant had not been extradited to the United States until long after the alleged tampering, and until he was extradited the government did not have a set of fingerprints known to be his. And no one has explained how fingerprints on another piece of material could have been transferred to the *adhesive* side of the tape, which was where they were found. It's one thing to press your finger on the adhesive side of a tape and remove the finger, leaving a print, but another thing to press a piece of paper containing your fingerprint on the adhesive side of the tape—try removing the paper without destroying the print.

[2] The defendant's petition, and amended petition, for rehearing did not defend the judge's conjecture that the weight discrepancy indicated tampering. We concluded that while the defendant could argue at trial that the jury should disregard the fingerprint evidence, there was no justification for excluding it in advance of trial on the “speculative” ground excogitated by the judge. Once the government presents evidence, as it did here (remember the ten witnesses), that adequate precautions had been taken to preserve the evidence challenged by the defendant, it has established admissibility, though at trial the defendant can challenge the adequacy of the precautions and present evidence of tampering. *United States v. Lee*, 502 F.3d 691, 697–98 (7th Cir.2007); *United States v. Kelly*, 14 F.3d 1169, 1175 (7th Cir.1994); *United States v. Brumfield*, 686 F.3d 960, 965 (8th Cir.2012); see

also *Melendez–Diaz v. Massachusetts*, 557 U.S. 305, 311 n. 1, 129 S.Ct. 2527, 174 L.Ed.2d 314 (2009). And that means by the way that even if our mandamus order was ultra vires it didn't undermine the fairness of the trial or the justice of the defendant's conviction. The fingerprint evidence should not have been excluded, and once admitted confirmed his guilt. We take up at the end of our opinion the defendant's distinct argument that the reassignment of the case to another judge prejudiced the jury, and show that that argument has no merit either.

The fresh issue relating to the fingerprint evidence is whether the prints of two fingers found on the adhesive tape were the defendant's. They were latent rather than patent fingerprints. Patent fingerprints are made by pressing a fingertip covered with ink on a white card or similar white surface, and are visible. Latent fingerprints are prints, usually invisible, left on a smooth surface when a person touches it with a finger or fingers. Laboratory techniques are employed to make a latent fingerprint visible so that it can be compared with other fingerprints. The latent prints on the adhesive tape on the bag of heroin in this case were found by a fingerprint examiner to match the defendant's patent prints made in the course of the criminal investigation, and the government therefore offered the match as evidence of the defendant's participation in the drug ring. The defendant argues that methods of matching latent prints with other latent prints or with patent prints have not been shown to be reliable enough to be admissible as evidence under the *484 standard for reliability set forth in *Fed.R.Evid.* 702, 703; *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 592–93, 113 S.Ct. 2786, 125 L.Ed.2d 469 (1993); and *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 149, 119 S.Ct. 1167, 143 L.Ed.2d 238 (1999).

The method the examiner used is called ACE–V and is the standard method for determining whether two fingerprints are from the same person. See Scientific Working Group on Friction Ridge Analysis, Study and Technology, “Standards for Examining Friction Ridge Impressions and Resulting Conclusions,” Sept. 13, 2011, www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf (visited Jan. 4, 2013); Michele Triplett & Lauren Cooney, “The Etiology of ACE–V and Its Proper Use: An Exploration of the Relationship Between ACE–V and the Scientific Method of Hypothesis Testing,” 56 *J. Forensic Identification* 345, 346 (2006). The defendant is therefore mounting a frontal assault on the use of fingerprint evidence in

litigation, an attack the courts have frequently rebuffed. See, e.g., *United States v. Havvard*, 260 F.3d 597, 601 (7th Cir.2001); *United States v. George*, 363 F.3d 666, 672–73 (7th Cir.2004); *United States v. Crisp*, 324 F.3d 261, 268–70 (4th Cir.2003); *United States v. Mitchell*, 365 F.3d 215, 235–46 (3d Cir.2004).

ACE–V is an acronym for analysis, comparison, evaluation, and verification, and has been described as follows:

The process begins with the analysis of the unknown friction ridge print (now often a digital image of a latent print). Many factors affect the quality and quantity of detail in the latent print and also introduce variability in the resulting impression.... If the examiner deems that there is sufficient detail in the latent print (and the known prints), the comparison of the latent print to the known prints begins.

Visual comparison consists of discerning, visually “measuring,” and comparing—within the comparable areas of the latent print and the known prints—the details that correspond. The amount of friction ridge detail available for this step depends on the clarity of the two impressions. The details observed might include the overall shape of the latent print, anatomical aspects, ridge flows, ridge counts, shape of the core, delta location and shape, lengths of the ridges, minutia location and type, thickness of the ridges and furrows, shapes of the ridges, pore position, crease patterns and shapes, scar shapes, and temporary feature shapes (e.g., a wart).

At the completion of the comparison, the examiner performs an evaluation of the agreement of the friction ridge formations in the two prints and evaluates the sufficiency of the detail present to establish an identification (source determination). Source determination is made when the examiner concludes, based on his or her experience, that sufficient quantity and quality of friction ridge detail is in agreement between the latent print and the known print. Source exclusion is made when the process indicates sufficient disagreement between the latent print and known print. If neither an identification nor an exclusion can be reached, the result of the comparison is inconclusive. Verification occurs when another qualified examiner repeats the observations and comes to the same conclusion, although the second examiner may be aware of the conclusion of the first.

National Research Council of the National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* 137–38 (2009).

***485** The methodology requires recognizing and categorizing scores of distinctive features in the prints, see Davide Maltoni et al., *Handbook of Fingerprint Recognition* 97–101 (2d ed. 2009); Federal Bureau of Investigation, *The Science of Fingerprints: Classification and Uses* 5–86 (2006), and it is the distinctiveness of these features, rather than the ACE–V method itself, that enables expert fingerprint examiners to match fingerprints with a high degree of confidence. That's not to say that fingerprint matching (especially when it involves latent fingerprints, as in this case) is as reliable as DNA evidence, for example. Forensic DNA analysis involves comparing a strand of DNA (the genetic code) from the suspect with a strand of DNA found at the crime scene. The comparison is done with scientific instruments and determines whether the segments are chemically identical. Errors are vanishingly rare provided that the strands of code are reasonably intact. As we explained in *United States v. Ford*, 683 F.3d 761, 768 (7th Cir.2012),

What is involved, very simply, in forensic DNA analysis is comparing a strand of DNA (the genetic code) from the suspect with a strand of DNA found at the crime scene. See “DNA Profiling,” *Wikipedia*, http://en.wikipedia.org/wiki/DNA_profiling (visited May 31, 2012). Comparisons are made at various locations on each strand. At each location there is an allele (a unique gene form). In one location, for example, the probability of a person's having a particular allele might be 7 percent, and in another 10 percent. Suppose that the suspect's DNA and the DNA at the crime scene contained the same alleles at each of the two locations. The probability that the DNA was someone else's would be 7 percent if the comparison were confined to the first location, but only .7 percent (7 percent of 10 percent) if the comparison were expanded to two locations, because the probabilities are independent. Suppose

identical alleles were found at 10 locations, which is what happened in this case; the probability that two persons would have so many identical alleles, a probability that can be computed by multiplying together the probabilities of an identical allele at each location, becomes infinitesimally small—in fact 1 in 29 trillion, provided no other comparisons reveal that the alleles at the same location on the two strands of DNA are different. This is the same procedure used for determining the probability that a perfectly balanced coin flipped 10 times in a row will come up heads all 10 times. The probability is $.5^{10}$, which is less than 1 in 1000.

Chemical tests can determine whether two alleles are identical, but a fingerprint analyst must visually recognize and classify the relevant details in the latent print—which is difficult if the print is incomplete or smudged. “[T]he assessment of latent prints from crime scenes is based largely on human interpretation.... [T]he process does not allow one to stipulate specific measurements in advance, as is done for a DNA analysis. Moreover, a small stretching of distance between two fingerprint features, or a twisting of angles, can result from either a difference between the fingers that left the prints or from distortions from the impression process.” National Research Council, *supra*, at 139.

Matching latent fingerprints is thus a bit like an opinion offered by an art expert asked whether an unsigned painting was painted by the known painter of another painting; he makes or rejects a match on the basis of visual evidence. Eyewitness evidence is similar. The eyewitness saw the perpetrator of a crime. His recollection of the perpetrator's appearance is analogous to a latent fingerprint. He sees the defendant at the trial—that sighting is *486 analogous to a patent fingerprint. He is asked to match his recollection against the courtroom sighting—and he is allowed to testify that the defendant *is* the perpetrator, not just that there is a close resemblance. A lineup, whether photo or in-person, is a related method of adducing matching evidence, as is handwriting evidence.

Matching evidence of the kinds that we've just described, including fingerprint evidence, is less rigorous than the kind

of scientific matching involved in DNA evidence; eyewitness evidence is not scientific at all. But no one thinks that only scientific evidence may be used to convict or acquit a defendant. The increasingly well documented fallibility of eyewitness testimony, see Elizabeth F. Loftus *et al.*, *Eyewitness Testimony: Civil and Criminal* (4th ed. 2007); *United States v. Ford*, *supra*, 683 F.3d at 764–66, has not banished it from criminal trials. *Perry v. New Hampshire*, — U.S. —, 132 S.Ct. 716, 728, 181 L.Ed.2d 694 (2012).

Evidence doesn't have to be infallible to be probative. Probability of guilt is a function of all the evidence in a case, and if items of evidence are independent of one another in the sense that the truth of any one item is not influenced by the truth of any other, the probability of guilt may be much higher if there is evidence from many independent sources (several eyewitnesses, an eyewitness plus fingerprints, etc.) than it would be were there only the evidence of one eyewitness, say. If “the prosecution submits three items of evidence of the defendant's guilt (and the defendant submits no evidence of his innocence), and the probability that item 1 is spurious is 10 percent, the probability that item 2 is spurious is also 10 percent, and likewise item 3 [, then the] probability that all three are spurious (assuming that the probabilities are independent—that is, that the probability that one piece of evidence is spurious does not affect the probability that another is), and therefore that the defendant should be acquitted, is only one in a thousand (.1 x .1 x .1).” *United States v. Williams*, 698 F.3d 374, 379 (7th Cir.2012).

[3] The defendant intimates that any evidence that requires the sponsorship of an expert witness, as fingerprint evidence does, must be found to be good science before it can be admitted under the doctrine of the *Daubert* case and Rules 702 or 703 of the Federal Rules of Evidence. But expert evidence is not limited to “scientific” evidence, however such evidence might be defined. *Kumho Tire Co. v. Carmichael*, *supra*, 526 U.S. at 150–51, 119 S.Ct. 1167; *Tuf Racing Products, Inc. v. American Suzuki Motor Corp.*, 223 F.3d 585, 591 (7th Cir.2000). It includes any evidence created or validated by expert methods and presented by an expert witness that is shown to be reliable. In a case involving an alleged forgery of a painting, there might be expert scientific evidence based on tests of the age of the canvas or paint; but there might also be expert evidence, offered by a dealer or art historian or other art expert, on the style of a particular artist. That evidence would be the expert's opinion, based on comparison with other paintings, of the genuineness of the painting alleged to be a forgery. See, e.g., *Levin v. Dalva Brothers, Inc.*, 459 F.3d

68, 78–79 (1st Cir.2006); *United States v. Tobin*, 576 F.2d 687, 690–91, 693 (5th Cir.1978).

Fingerprint experts such as the government's witness in this case—who has been certified as a latent print examiner by the International Association for Identification, the foremost international fingerprint organization (there are only about 840 IAI-certified latent examiners in the world, out of 15,000 total examiners)—receive extensive *487 training; and errors in fingerprint matching by expert examiners appear to be very rare. Of the first 194 prisoners in the United States exonerated by DNA evidence, none had been convicted on the basis of erroneous fingerprint matches, whereas 75 percent had been convicted on the basis of mistaken eyewitness identification. Greg Hampikian *et al.*, “The Genetics of Innocence: Analysis of 194 U.S. DNA Exonerations,” 12 *Annual Rev. of Genomics and Human Genetics* 97, 106 (2011). The probability of two people in the world having identical fingerprints is not known, but it appears to be extremely low. Steven M. Stigler, “Galton and Identification by Fingerprints,” 140 *Genetics* 857, 858 (1995); David A. Stoney & John I. Thornton, “A Critical Analysis of Quantitative Fingerprint Individuality Models,” 31 *J. of Forensic Sciences* 1187 (1986). The great statistician Francis Galton estimated the probability as 1 in 64 billion. Galton, *Finger Prints* 110 (1892); Stigler, *supra* at 858. That was not an estimate of the probability of a mistaken matching of a latent to a patent or another latent fingerprint. Yet errors in such matching appear to be very rare, though the matching process is judgmental rather than scientifically rigorous because it depends on how readable the latent fingerprint is and also on how distorted a version of the person's patent fingerprint it is. Examiners' training includes instruction on how to determine whether a latent print contains enough detail to enable a reliable matching to another print. Ultimately the matching depends on “subjective judgments by the examiner,” National Research Council, *supra*, at 139, but responsible fingerprint matching is admissible evidence, in general and in this case.

The other issues presented by the appeal that merit discussion arise from the interruption of the trial by the mandamus proceeding and the resulting reassignment of the case to a different district judge. The consequence was an eleven-day hiatus in the trial. The defendant argues that when the trial resumed, the jurors, remembering the skeptical remarks that the original judge had made about the government's evidence, must have thought that he had been “punished” for siding with the defendant by being removed and therefore that the jury

should convict. That is unpersuasive conjecture. Because of sickness most commonly, but sometimes for other reasons, such as belated discovery of a ground for recusal, a judge is sometimes replaced during a trial and when that happens the new judge tells the jury that such replacements happen occasionally and the jurors are not to worry about the change in judges or speculate about the reason for it. The new judge in this case didn't explain the cause of the delays but did say:

It is very important for me to emphasize this instruction, that however you may feel about the delays in this case, you are not to hold those feelings against anybody in this courtroom.... In fact, I am going to instruct you right now that you not speculate about the causes or reasons for the delays at all.... To the extent that you have been told or you have come to believe that the delays are somehow the fault of the government or the fault of the defense counsel, I am instructing you that you put those concerns out of your mind completely.... At the end of this case, we will not be asking you, did the trial go smoothly? And if not, whose fault was it? That will not be a question you will be asked to consider. The only question you will be asked to consider at the conclusion of this case is, did the government meet its burden of proof? That's the only question. And concerns about delays are not to be in your mind at all.... From time to time there are reasons that we have to interrupt the smooth progress of a *488 trial. It's happened to me before. This was one of those occasions.... Your consideration of the evidence should not be influenced in any way by any assumptions you may have made or any conclusions you may have drawn about delays.

There is no history of which we're aware of miscarriages of justice resulting because juries draw erroneous inferences from the replacement of a judge. See *United States v. Gayles*, 1 F.3d 735, 738 (8th Cir.1993); *United States v. LaSorsa*, 480 F.2d 522, 531 (2d Cir.1973).

[4] The defendant complains that the new judge pressured the jury to complete its deliberations in a day and that with more time it might have acquitted him. There is no evidence to support that accusation of a judge noted for her patience. The first judge had assured the jury that the trial would not interfere with any of the jurors' vacation schedules. When trial resumed on August 2 the jury was down to 12 because one of the two alternates had been excused and the other had replaced a juror who had been excused. One of the remaining jurors had long-standing vacation plans for August 5, and the original judge had (with the government's consent) assured her when the government sought mandamus and the trial was adjourned that she would not need to show up on or after that date. When the trial resumed, another juror asked in open court what the jury should do in light of the possibility that the juror with vacation plans would leave before the trial ended. In response, and without objection by the defendant's lawyer, the judge said "we can't proceed" with fewer than 12 jurors. That was true (since the parties would not stipulate to a jury of 11, see [Fed.R.Crim.P. 23\(b\)\(2\)](#)), though what was also true but she rightly did not say, because it would have sown confusion, is that while the *trial* could not continue without 12 jurors, if once the jury retired for its deliberations one of the jurors then decamped the judge could allow the remaining 11 to render a verdict even without the lawyers' consent. [Fed.R.Crim.P. 23\(b\)\(3\)](#).

August 4 turned out to be the last day of the trial. Closing arguments and the reading of the instructions to the jury took until the afternoon. The jury retired to consider its verdict at about 3:45 and returned 7 hours later with a verdict of guilty on eight counts and not guilty on the remaining six. The defendant argues that the jurors had rushed to complete their deliberations, knowing there would not be 12 jurors the next day. Given the strength of the government's case and the length of the jury's deliberations, and the fact that there was only one defendant and that the jury acquitted him on some counts, it is unlikely that even if they hadn't been expecting to lose the twelfth juror the next day, the jurors would have taken more time to deliberate than they did, though they

might have broken at dinner time and resumed the following morning. The judge did not, as in the cases that the defendant cites to us, [United States v. Blitch](#), 622 F.3d 658, 670 (7th Cir.2010), and [United States v. Chaney](#), 559 F.2d 1094, 1098 (7th Cir.1977), set a deadline, either explicit or implicit, for the jury's deliberations. On the contrary, after instructing the jury, and only moments before the jury left the courtroom to deliberate, the judge told them: "I think I mentioned earlier that from this point on, the schedule is up to you. I realize [by] the way that the trial has been bumpy, and I will make every effort to accommodate your schedule from this moment on, whatever your decisions are. I appreciate your time. I think all of us do. You are excused to deliberate on your verdict." That was the opposite of pressuring the jury to complete its deliberations in a day. The jurors were unlikely to feel rushed when the judge had gone out of her way to tell them that she would make every effort to *489 accommodate their schedules. Had the jurors been unable to agree on a verdict on August 4, the foreman would have told the judge that they couldn't reach a verdict and she would have either discharged them and declared a mistrial or allowed the 11 remaining jurors to return the next day and deliberate.

When the jury retired to deliberate, knowing that one juror would leave on vacation the next day and perhaps believing that 12 jurors had to be present to render a verdict, no juror asked the judge a question such as: "Does this mean we must render a verdict by the end of the day or can we just report our inability to reach a verdict?" Or: "What if we can't complete our deliberations by the end of the day?" Such questions would have flagged concerns that the judge would doubtless have addressed. No questions were asked. That suggests that the jurors were not concerned that the trial might end without a verdict unless they rushed their deliberations.

AFFIRMED.

All Citations

704 F.3d 480

FORENSICS

CRIMINAL LAW

The sea change in our understanding of fingerprint identification testimony, and of its very real limits, requires lawyers and courts to look beyond the inadequate guidance past judicial decisions have provided, says attorney Gabriel A. Fuentes in this BNA Insight.

The author examines the fingerprint identification landscape, and concludes when *Daubert* or *Frye* is appropriately applied, there should be no place in the courtroom for fingerprint examiners to make individualization claims, at least until research exists to back up the validity of an opinion that will be so compelling to jurors who have been told for their entire lives that fingerprint identification is infallible.

Toward a More Critical Application of *Daubert* in Criminal Cases: Fingerprint Opinion Testimony After the National Academy of Sciences Report

method has been shown to establish—consistently and



BY GABRIEL A. FUENTES

More than three years have passed since the National Academy of Sciences released its groundbreaking report on the state of the forensic sciences in the United States. The NAS stated bluntly that with the exception of nuclear DNA analysis, no forensic

reliably—a connection between trace evidence and a specific individual.¹

The NAS pointed out that fingerprint identification in particular lacked any genuine research to validate claims that examiners regularly made before courts over the years, including claims that examiners may “individualize” a latent fingerprint to a single person.² Even as the fingerprint examiner community is backing away from those claims, courts are continuing to allow examiners to make them in criminal cases. That should stop, at least until prosecutors can offer adequate evidence backing the validity of this opinion that has increasingly been called into doubt after the 2009 NAS report.

The NAS report and its aftermath should prompt prosecutors, defense lawyers, and judges to make a stark re-examination of how fingerprint identification evidence is evaluated, presented, and admitted in criminal trials. Criminal practitioners are in a new era, in which we should demand that courts give fingerprint identification greater scrutiny to ensure that examiners’ claims, and particularly those which contend latent prints may be “individualized” to a single person, are backed by research indicating that the claims are valid. Past judicial acceptance of the testimony must no longer be sufficient.

The Fingerprint Identification Landscape

One of the most persuasive aspects of fingerprint identification is its longevity in the public mind. Generations of jurors have grown up around the assumption that everyone’s fingerprint is unique, and that by a comparison of a latent fingerprint left at a crime scene, an examiner can identify the person whose finger left the latent print. In Mark Twain’s 1894 novel *Pudd’nhead Wilson*, the main character, a lawyer, proves his client innocent of a murder by arguing that a bloody fingerprint on the knife belonged to someone else, and even some of the early judicial decisions accepting fingerprint analysis cited Twain’s novel as among the relevant authorities.³

Historically, the proponents of fingerprint analysis have pointed out that fingerprints have been used as seals or in lieu of signatures “since antiquity,” but that the first U.S. court decision approving the admission of fingerprint identification as evidence in a criminal matter was *Illinois v. Jennings*.⁴ Remarkably, though, the *Jennings* court relied on an entry in the 1911 *Encyclopedia Britannica* for the notion that fingerprint identifi-

cation had validity.⁵ For its research basis for validity of fingerprint identification, the *Encyclopedia* relied on the 1892 writings of Sir Francis Galton.⁶

Galton’s signature work, *Fingerprints*, was not based on any research into the uniqueness of individual fingerprints or into the ability of anyone to individualize them.⁷ Galton’s work was nonetheless cited by the same early 20th Century U.S. courts which had relied on Twain’s *Pudd’nhead Wilson* to allow fingerprint evidence, and the cited passage was Galton’s supposition that if only the Biblical contemporaries of Jezebel had her known prints on file, they could have identified her after the dogs of Jezreel devoured all but her skull, hands and feet.⁸

Jennings and its reliance on the *Encyclopedia* (and, by extension, Galton) would be a quirky anachronism if not for the fact that *Jennings* was among the authorities cited by two of the leading federal appellate decisions today supporting the admission of fingerprint identification evidence based on the long history of its judicial acceptance.⁹ These two pre-NAS Report cases, *United States v. Havvard* and *United States v. Crisp*, represent the longstanding judicial view that fingerprint identification evidence should be admitted because it has always been admitted.

In both of these cases, the federal courts of appeal affirmed the admission of examiner testimony that the defendant himself was the exclusive source of the latent fingerprint.¹⁰ The U.S. Department of Justice continues to cite *Havvard* and *Crisp* today in support of the proposition that fingerprint examiners should be allowed to testify that a latent print is a “match” to a particular individual.¹¹

The ACE-V Method and the NAS Report

Since 1959, the identification method employed by today’s examiners has gone by the name of ACE-V, an acronym for the words “analysis, comparison, evaluation, and verification [by a second examiner].”¹² Much

⁵ *Jennings*, 96 N.E. 1081.

⁶ 10 *Encyclopedia Britannica* 376 (Cambridge Univ. Press 11th ed. 1911).

⁷ Francis Galton, *Fingerprints* passim (De Capo Press ed. 1965) (“Galton”).

⁸ See *supra* n. 4; Galton at 113. The passage, alluding to the Second Book of Kings, is as follows: “We read of the dead body of Jezebel being devoured by the dogs of Jezreel, so that no man might say, ‘This is Jezebel,’ and that the dogs left only her skull, the palms of her hands, and the soles of her feet; but the palms of the hands and the soles of the feet are the very remains by which a corpse might be most surely identified, if impressions of them, made during life, were available.”

⁹ *United States v. Crisp*, 324 F.3d 261, 266 (4th Cir. 2003); *United States v. Havvard*, 260 F.3d 597, 601 (7th Cir. 2001).

¹⁰ See *Havvard*, 260 F.3d at 598 (“A latent fingerprint lifted from one of the handguns was later matched to an exemplar fingerprint obtained from Havvard.”); *Crisp*, 324 F.3d at 265 (“a fingerprint expert . . . testified that Crisp’s right palm had produced a latent print that had subsequently recovered” from a handwritten note used in a robbery).

¹¹ Brief of the United States at 32-37, *United States v. Clacy Watson Herrera*, No. 11-2894 (7th Cir. April 11, 2012). The author is counsel of record in this pending appeal.

¹² NAS Report at 137. The “analysis” phase refers to the analysis of the latent fingerprint impression to assess its ridge formations and the clarity of those formations, taking account of factors such as pressure, distortions, and the medium from

¹ National Research Council, National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* 7 (National Academies Press 2009) (“NAS Report”).

² *Id.* at 43-44, 143-45.

³ *State v. Kuhl*, 175 P. 190, 191 (Nev. 1913); *Stacy v. State*, 292 P. 885, 886 (Okla. Crim. App. 1930).

⁴ 96 N.E. 1077 (Ill. 1911). See Andre A. Moenssens, Fred E. Inbau, and James E. Starrs, *Scientific Evidence in Criminal Cases* 416, 438 (3d ed. 1986). The authors noted: “The *Jennings* case further held that persons experienced in the matter of fingerprint identification may give their opinions as to whether or not the fingerprints found at the scene of a crime corresponded with those of the accused.” *Id.* at 438. Thus, fingerprint identification started with the bold claim that the examiner’s ability to achieve “individualization” was beyond question.

of the NAS committee's work revolved around assessing the validity and reliability of the ACE-V method of fingerprint identification. In the end, the NAS Report concluded that ACE-V "provides a broadly stated framework for conducting friction ridge analyses," but added:

However, this framework is not specific enough to qualify as a validated method for this type of analysis. ACE-V does not guard against bias; is too broad to ensure repeatability and transparency; and does not guarantee that two analysts following it will obtain the same results. For these reasons, merely following the steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results.¹³

The NAS Report made numerous other significant observations about the degree of validity and reliability underpinning the ACE-V method and fingerprint analysis generally:

- A "thorough analysis of the ACE-V method" by experts Ralph and Lyn Haber concluded unambiguously: "We have reviewed available scientific evidence of the validity of the ACE-V method and found none."¹⁴

- Fingerprint examiners are "unjustified" in claiming they can match a latent fingerprint to a particular finger to the exclusion of all others in the world.¹⁵

- Claims of "uniqueness" in friction ridge patterns are unsubstantiated, in that "[t]he determination of uniqueness requires measurements of object attributes, data collected on the population frequency of variation in those attributes, testing of attribute independence, and calculations of the probability that different objects share a common set of observable attributes." While "[s]ome scientific evidence support[s] the presumption that friction ridge patterns are unique to each person and persist unchanged throughout a lifetime. . . . None of these variabilities—of features across a population of fingers or of repeated impressions left by the same

which the impression was recovered. Christophe Champod, Chris Lennard, Pierre Margot, and Milutin Stoilovic, *Fingerprints and Other Ridge Skin Impressions* 15 (CRC Press LLC 2004). "Comparison" is the actual comparison of the features in the latent impression to those in a known or exemplar print, focusing on the different types of features (broken down into levels 1, 2, and 3) and accounting for the tolerances dictated by the quality of the latent impression. *Id.* at 21. "Evaluation" refers to the drawing of an inference about the identity of the source of the latent impression; similarities are evaluated and quantified, and dissimilarities are studied for whether they lead to an exclusion, or whether they may be explained consistent with an identification. *Id.* at 21-24. "Verification" refers to a second examiner's review of the first examiner's conclusions. *Id.* at 39-40. Champod, a respected fingerprint examiner, has acknowledged that as a protocol, the ACE-V method "does not completely fulfill the requirements . . . of (a) a fully articulated descriptive model, (b) a detailed and systematic account of the variation of the features, and (c) a transparent decision model." *Id.*, preface; see also Jennifer L. Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 *Brook. L. Rev.* 1209, 1268 (2010) ("ACE-V is extremely vague, and does not come close to providing a fully developed and adequately articulated method with detailed specifications.").

¹³ NAS Report at 142.

¹⁴ *Id.* at 143.

¹⁵ *Id.* at 142.

finger—has been characterized, quantified, or compared."¹⁶

Co-chaired by the Hon. Harry T. Edwards of the D.C. Circuit Court of Appeals and Brown University biostatistics professor Constantine Gatsonis, the NAS's Committee on Identifying the Needs of the Forensic Science Community included forensic scientists and legal and technical experts, and it heard presentations from a wider range of experts including fingerprint examiners and crime lab directors.¹⁷ After the report was issued, Judge Edwards commented that the NAS project had educated him about how the forensic disciplines as a whole were not as well-grounded in validity and reliability as he had assumed.¹⁸ On fingerprints specifically, Judge Edwards told a PBS documentary in April 2012 that when partial or smudged latent fingerprints present examiners with "hard cases," the examiners did not have research to back their claims that they could make identifications, and moreover, courts had been "misled" when examiners told them that the identification method had a zero error rate.¹⁹

Edwards added:

You had judges who were absolutely naïve and ill-informed writing in these opinions. The first case in which it came up was the FBI expert told [another circuit court of appeals] . . . zero error rate. Then another circuit court picked up the same information and said it's a zero error rate That's repeated over and over again without us ever understanding that's completely inaccurate, and you don't have the science to support it.²⁰

Havvard and Crisp: A Now Outmoded View

Judge Edwards did not specify which circuit courts of appeal he was referencing, but notably, *Havvard* and then *Crisp* relied on then-uncontradicted FBI testimony (given before the district court in *Havvard* and cited by the Fourth Circuit later in *Crisp*) that fingerprint identification had a zero error rate.²¹ Aside from the NAS Report's dismissal of the notion of zero error rates in fingerprint analysis as not realistic,²² prominent organizations from the fingerprint examiner community have also acknowledged that fingerprint analysis simply does not have a zero error rate. Shortly after the NAS Report was released, the International Association for Identification (IAI) and the Scientific Working Group on Friction Ridge Analysis, Study, and Technology (SWGFAST) issued statements backing away from the

¹⁶ *Id.* at 142-44.

¹⁷ *Id.* at v, xi-xii.

¹⁸ Judge Edwards stated:

I started the NAS project with no skepticism regarding the forensic science community. Rather, I assumed, as I suspect many of my judicial colleagues do, that the forensic disciplines are well grounded in scientific methodology and that crime laboratories and forensic practitioners follow proven practices that ensure the validity and reliability of forensic evidence offered in court. I was surprisingly mistaken in what I assumed.

The Hon. Harry T. Edwards, Co-Chairman, NAS Committee on Identifying the Needs of the Forensic Science Community, Presentation at the Superior Court of the District of Columbia, Conference on The Role of the Court in an Age of Developing Science & Technology (May 6, 2010).

¹⁹ Interview by PBS Frontline with the Hon. Harry T. Edwards (released April 16, 2012).

²⁰ *Id.*

²¹ *Havvard*, 260 F.3d at 599; *Crisp*, 324 F.3d at 269.

²² NAS Report at 143.

idea that a zero error rate was plausible or appropriate for court testimony.²³

But the implausibility of a zero error rate, and its rejection by the NAS Report and the fingerprint examiner community, is only the beginning of the problems underlying *Crisp* and *Havvard* as standard-bearing cases for the practical and legal limits on a fingerprint examiner's testimony about identification. Both of these federal appellate decisions affirmed the admission of fingerprint examiner testimony to a "match" between the latent print and the defendant, based heavily on past judicial acceptance of the "science" of fingerprinting.

In *Havvard*, the Seventh Circuit noted that the district court had relied on the "100 years of successful use" of fingerprint analysis in criminal trials.²⁴ The Seventh Circuit then spoke approvingly of the district court's conclusion that "fingerprinting techniques have been tested in the adversarial system."²⁵

In *Crisp*, the Fourth Circuit cited the foregoing language from *Havvard* for the conclusion 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."²⁶

The courts in *Havvard* and *Crisp* were considering challenges to the evidence under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,²⁷ in which the Supreme Court designated the trial courts as gatekeepers to confirm the validity and reliability of scientific evidence offered through experts, and *Kumho Tire Co. v. Carmichael*,²⁸ in which the Court extended the *Daubert* framework to all expert testimony involving technical or specialized knowledge.

The core of the *Daubert* inquiry surrounded whether the proposed testimony is supported by "appropriate validation—i.e., 'good grounds,' based on what is known."²⁹ *Daubert* set forth a set of flexible factors for making that sort of determination, including whether the knowledge "can be (and has been) tested."³⁰

The conclusions in *Crisp* and *Havvard* that 100 years of "adversarial testing" supported admissibility under *Daubert*'s notion of "testing" have prompted some withering criticism. In a passage quoted in the NAS Report, one commentator called this proposition "a silly statement" that resulted from judges being so reluctant to exclude fingerprint evidence that the prospect of doing so "stilled their critical faculties."³¹

The NAS committee called this comment "a telling critique, especially when one compares the judicial de-

isions that have pursued rigorous scrutiny of DNA typing with decisions that have applied less stringent standards of review in cases involving fingerprint evidence."³² In criminal cases, there is much reason to believe that the reliability of fingerprint analysis and identification methods, including ACE-V, was rarely if ever "tested," because criminal defense lawyers often lack the resources or skill to mount a meaningful *Daubert* challenge.³³

Instead, the lack of meaningful *Daubert* challenges led to a "vicious cycle" in which the long history of courts unquestioningly admitting the evidence without challenge contributed to an unwillingness by advocates to make *Daubert* challenges.³⁴ But even assuming that "testing" in the adversarial system represents some form of "testing," it is not the sort of "testing" contemplated by *Daubert*.

Daubert's central premise that judges should act as gatekeepers by erecting a heightened reliability screen for expert evidence assumes that the ordinary adversarial processes of a case will not be sufficient to ensure the validity and reliability of such evidence.³⁵ If past "adversarial testing" in the courts is sufficient to meet *Daubert*'s heightened reliability screen, then a discipline such as fingerprint identification faces no heightened screen at all. The gatekeeping role of the district court, under *Daubert*, would simply be subordinated to the decisions of past courts. Yet this very result appears to have occurred under *Havvard* and *Crisp*, in the absence of validation studies from the records in those cases.

Moreover, the courts in *Havvard* and *Crisp* decided those cases based on the records that were then before them. The state of the record in those pre-NAS Report cases is extremely thin. In *Havvard*, the defendant offered a 2000 National Institute of Justice solicitation for proposals for more fingerprint research studies, along with a 1995 Collaborative Testing Service report concerning fingerprint examination error rates—but the Seventh Circuit noted that neither of those documents was in the district court record, so neither could be considered on appeal.³⁶ Practically, then, the Seventh Circuit had virtually no study or other meaningful record before it in *Havvard* to show that fingerprint identification or ACE-V lacked demonstrated validity.

In *Crisp*, the record included the 2000 NIST solicitation and a few pre-NAS Report law review articles, but, as the Fourth Circuit noted, "no studies demonstrating the unreliability of fingerprinting analysis."³⁷ Today, many highly relevant reports, articles, and other materials are widely known and available to any court considering a *Daubert* challenge to fingerprint identification, including but not limited to:

- the NAS Report;

²³ Press Release, Robert J. Garrett, President, International Association for Identification (February 19, 2009) (suggesting to IAI members that they not assert 100% infallibility or zero error rate in testimony about fingerprint comparisons); SWGFAST Position Statement (August 3, 2009) ("SWGFAST acknowledges that errors do occur and furthermore that claims of zero error rate in the discipline are not scientifically plausible.")

²⁴ *Havvard*, 260 F.3d at 601.

²⁵ *Id.*

²⁶ *Crisp*, 324 F.3d at 268.

²⁷ 509 U.S. 579 (1993).

²⁸ 526 U.S. 137 (1999).

²⁹ *Daubert*, 509 U.S. at 590.

³⁰ *Id.* at 593.

³¹ NAS Report at 103, citing 1 D.L. Faigman, M.J. Saks, J. Sanders, and E.K. Cheng, *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 1:1, p. 4 (Thomson/West 2007-2008).

³² NAS Report at 103.

³³ Nancy Gertner, "National Academy of Sciences Report: A Challenge to the Courts," *American Bar Association, Criminal Justice* 11 (Spring 2012); Peter J. Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice: And Some Suggestions for Reform*, 95 Am. J. Pub. Health S107, S110 (2005).

³⁴ Nancy Gertner, *Commentary on the Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. Rev. 789, 790 (2011).

³⁵ Mnookin, *supra* n. 13, 75 Brook. L. Rev. at 1244.

³⁶ *Havvard*, 260 F.3d at 600-01.

³⁷ *Crisp*, 324 F.3d at 267.

- the 2006 report of the FBI Inspector General concerning the widely publicized fingerprint misidentification of Oregon lawyer Brandon Mayfield in the 2004 Madrid train bombing;³⁸

- the 2011 report commissioned by the British government on a widely publicized (in the United Kingdom) fingerprint misidentification case involving Scottish police officer Shirley McKie;³⁹

- a 2012 study by the NIST proposing a series of systematic and process improvements for the examination of latent fingerprints;⁴⁰

- recent re-evaluations, by prominent organizations in the fingerprint examiner community, of the limits of fingerprint identification;⁴¹ and

- a growing consensus of academics and professionals (including forensic scientists) that the forensic sciences in the United States need to be backed by a stronger research culture.⁴²

Finally, the courts in *Havvard* and *Crisp* suffered not only from the inadequacy of the record in those cases, and from the then-inadequately developed state of public knowledge concerning the state of the research base underlying fingerprint identification, but also from a lack of focus on how the relevant questions ought to have been framed before those courts. Again, in both cases, the government sought to introduce an opinion that a latent fingerprint was a “match” to the defendant, or that the defendant was the source of the latent fingerprint. The only way to interpret such testimony, if it is offered without limitation, is to consider it as an assertion that the latent fingerprints belong to the defendant only, to the exclusion of all others in the world.

Yet neither *Havvard* nor *Crisp* seemed to grasp that this particular form of testimony might require more rigorous vetting under *Daubert*. *Havvard* spoke of “the issue of reliability of fingerprint evidence” and whether “fingerprint comparisons are not reliable.”⁴³

Crisp described the defendant’s position as an argument for “the wholesale exclusion of a long-accepted form of expert evidence. Such a drastic step is not required of us under *Daubert*, and we decline to take it.”⁴⁴ But the question need not have been as “drastic.” The proper question was whether the examiners in

those two cases could validly and reliably offer testimony that the latent fingerprints they examined were indeed those of the defendant and no other.

The *Daubert* analysis of the ability of those examiners to give that testimony should have involved a searching inquiry into what research supported the idea that individualization was grounded in good science, or, if the basis was merely experiential, what experience, training, or other basis could give the courts a reason to believe such an opinion could have validity in the first instance.

Instead, *Havvard* and *Crisp* framed the issue much more broadly, and in a way that suggested far more dire consequences to fingerprint identification than even the NAS Report would suggest years later. Not surprisingly, the courts in those cases were reluctant to throw out a century of jurisprudence on fingerprint analysis. *Daubert* never required them to go that far, and nor does it today require a court to hold that fingerprint analysis is “unreliable” or generally inadmissible. *Daubert* does require that if the specific testimony being offered has not been or cannot be shown to be reliable, the district court as gatekeeper should not admit it.⁴⁵ The practical consequence of this sort of judicial approach is simply a heightened reliability screen for expert evidence, and that is what the Supreme Court had in mind all along in *Daubert*. The consequence is not that fingerprint identification testimony would be excluded under all circumstances and in all its forms.

Accordingly, given the significant differences in what we now know about fingerprints versus what the *Havvard* and *Crisp* courts knew when those cases were decided, and given the inappropriate way in which the relevant *Daubert* issues were framed in both of those cases, the *Havvard* and *Crisp* cases are of little use today in considering the reliability of the recurring claim by examiners that they can individualize or “match” a latent fingerprint to a specific individual person.

From Here Forward: A Need for Critical Judicial Analysis

Three years after the NAS Report, it is time to acknowledge that the report did bring about changes in how fingerprint identification evidence should be viewed by courts. Before the NAS Report, the lack of demonstrated validity and reliability of the ACE-V method was not widely known. As Judge Edwards has noted, courts were misled into believing that fingerprint identification had a zero error rate.⁴⁶ Courts were also told that experts could testify to a match between evidence and an individual source, but as Judge Edwards has noted, “that’s not accurate,” because the likelihood

³⁸ FBI Office of Inspector General, *A Review of the FBI’s Handling of the Brandon Mayfield Case* (March 2006);

³⁹ Sir Anthony Campbell, *The Fingerprint Inquiry Report* (APS Group Standard 2011).

⁴⁰ National Institute of Standards and Technology, National Institutes of Justice, *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach* (Feb. 2012) (“NIST Report”).

⁴¹ Position Statement. Scientific Working Group on Friction Ridge Analysis, Study, and Technology (April 21, 2012) (removing from SWGFAST’s definition of “individualization” the previously supposed ability of an examiner to match a latent fingerprint to one individual, to the exclusion of all others in the world).

⁴² Jennifer L. Mnookin, Simon A. Cole, Itiel E. Dror, Barry A.J. Fisher, Max M. Houck, Keith Inman, David H. Kaye, Jonathan J. Koehler, Glenn Langenburg, D. Michael Risinger, Norah Rudin, Jay Siegel, and David A. Stoney, *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. Rev. 725 (2012).

⁴³ *Havvard*, 260 F.3d at 600.

⁴⁴ *Crisp*, 324 F.3d at 268.

⁴⁵ Nor should such evidence be admissible in jurisdictions following *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), if the evidence is not “generally accepted” in the relevant scientific community. We can debate what constitutes general acceptance, and what constitutes the relevant scientific community, but the NAS Report coupled with the fingerprint examiner community’s retreat from zero error rate and individualization claims give *Frye* jurisdictions much to consider in evaluating proffered fingerprint identification testimony, particularly if that testimony makes individualization claims not now supported by the relevant community of fingerprint examiners and research scientists.

⁴⁶ See *supra* n. 19.

of a match should instead be expressed in probabilistic terms.⁴⁷

After the NAS Report, the 2012 NIST report recommending systemic improvements to latent fingerprint examination procedures specifically stated that testimony individualizing a latent print to one source, to the exclusion of all others, is “needlessly strong, not yet adequately supported by fundamental research, and impossible to validate solely on the basis of experience.”⁴⁸ The scientific working group of fingerprint examiners, SWGFAST, has rejected individualization, at least for now, as “not supported by research.”⁴⁹ So has the official report into the inquiry into the fingerprint misidentification of Scottish police officer Shirley McKie.⁵⁰

Yet reports of fingerprint examiners offering opinions that latent fingerprints are a “match” to a particular individual continue to persist.⁵¹ In at least three reported decisions since the summer of 2011, federal district courts have approved the admission of examiner testimony that latent fingerprints match or are those of the accused.⁵² These more recent cases follow a series

⁴⁷ *Id.*

⁴⁸ NIST Report at 197.

⁴⁹ Position Statement. Scientific Working Group on Friction Ridge Analysis, Study, and Technology (April 21, 2012).

⁵⁰ Sir Anthony Campbell, *The Fingerprint Inquiry Report* 682-83 (APS Group Standard 2011).

⁵¹ See *supra* n. 19.

⁵² Each of these three decisions has individual problems. In *United States v. Gutierrez-Castro*, 805 F. Supp. 2d 1218, 1228-35 (D. N.M. 2011), the court allowed an examiner to testify to a fingerprint match to the defendant but declined to “certify” the examiner as an expert. In finding that the proffered testimony was reliable, the court relied entirely on pre-NAS case law, most notably a conclusion in *United States v. Baines*, 573 F.3d 979 (10th Cir. 2009), in which the Tenth Circuit bought into the fallacy that fingerprint identification had been subject to adequate testing in the form of “the world of criminal investigations, court proceedings, and other practical applications, such as identification of victims of disasters.” *Gutierrez-Castro*, 805 F. Supp. 2d at 1230-31, quoting *Baines*, 573 F.3d at 990. Moreover, the *Gutierrez-Castro* court never differentiated between the reliability of fingerprint identification as a discipline and of individualization claims, which the discipline now has acknowledged are beyond its limits. See *supra* n. 49-50. Similarly, in *United States v. Stone*, No. 10-20123, 2012 BL 229856, at *3 (E.D. Mich. Jan. 25, 2012), the court allowed individualization testimony over a defense objection the court characterized as “a wholesale attack on latent fingerprint identification evidence in general,” as well as the holdings in *Crisp* and *Havvard*. *Stone* also relied on *United States v. Pena*, 586 F.3d 105 (1st Cir. 2009), another appellate matter argued before issuance of the NAS Report. Notably, the now-retired district judge who denied the *Daubert* challenge litigated on appeal in *Pena* wrote recently that the challenge was perfunctory, that no *Daubert* hearing took place because the defendant waived it, and that “[t]o my amazement, the defense announced it had no witnesses, no experts, and did not even wish to take advantage of the opportunity to examine the government witnesses. The motion, he insisted, had only been brought ‘for the record.’” Gertner, *supra* n. 34, at 10. Finally, in the most recent case as of this writing, a court rejected, without a hearing, a *pro se* *Daubert* challenge in which the *pro se* defendant declined to file a reply brief, and the district court accepted a magistrate judge’s recommendation that relied heavily on *Stone* and the same problematic roster of cases cited in *Stone*. See *United States v. Campbell*, No. 11-CR-460, 2012 BL 157136 (N.D. Ga. June 22, 2012), and *United States v. Campbell*, No. 11-CR-460, 2012 BL 257339, at *1 (N.D. Ga. April 19, 2012).

of post-NAS Report federal district court cases which similarly did not address individualization (likely because it was not clearly presented), or which largely deferred to earlier pre-NAS Report decisions that had held fingerprint identification in general to be reliable.⁵³ Meanwhile, prosecutors still argue that individualization is valid under *Daubert*, in light of precedents such as *Havvard* and *Crisp*.⁵⁴

Meanwhile, the notion of the infallibility of fingerprint identification is embedded deeply in the public mindset, fueled by more than a century of popular culture, from Mark Twain’s *Pudd’nhead Wilson* to Jack Webb’s *Dragnet* and beyond.⁵⁵

Conclusion

The sea change in our knowledge and understanding of fingerprint identification testimony, and of its very real limits, requires lawyers and courts to look beyond the inadequate guidance that past judicial decisions have provided. In particular, the recurring but now discredited claim that an examiner may individualize a latent fingerprint to a single human source must be excluded from the courtroom, at least until an adequate research basis can be advanced to support that claim.

This sort of careful judicial scrutiny would be a needed fresh start, and a needed departure from a line of decisions that continues to build upon itself, and upon the flawed roster of pre-NAS Report decisions that admitted fingerprint identification testimony, including individualization, for some 100 years without any demonstrated basis of reliability. Courts today must reserve their power to exclude, under *Daubert* and *Kumho Tire*, evidence that cannot be shown to be reliable or valid.⁵⁶

⁵³ These courts generally relied on mere acceptance of the method within the fingerprint examiner community, or on the notion that the NAS Report was not a proposal for “law reform,” and thus did not place significant weight in the absence of a true research basis for validity of the ACE-V method. *E.g.*, *United States v. Council*, 777 F. Supp. 2d 1006, 1011-12 (E.D. Va. 2011) (allowing evidence because “[f]riction ridge experts maintain widely recognized standards” which the expert applied to the prints at issue in this case); *United States v. Aman*, 748 F. Supp. 2d 531, 541 (E.D. Va. 2010) (ruling that important questions raised in the NAS Report, most significantly the subjectivity of examiners, are susceptible to vigorous attack on cross-examination); *United States v. Cerna*, No. CR 08-0730 WHA, 2010 BL 321214, at *8 (N.D. Cal. Sept. 1, 2010) (“the NAS report is not a law reform proposal and its findings are not coterminous with whether forensic evidence in a particular case is admissible”); and *United States v. Rose*, 672 F. Supp. 2d 723, 725 (D. Md. 2009) (“the Report itself did not conclude that fingerprint evidence was unreliable such as to render it inadmissible under Fed. R. Evid. 702”).

⁵⁴ See *supra* n. 12.

⁵⁵ See, *e.g.*, Jack Webb, *The Badge* 117 (Thunder Mouth’s Press ed. 2005 [first ed. 1958]) (referring to early Los Angeles Police Department criminalists as having taught detectives that “[n]othing in this world . . . is found in exact duplication—not even two billiard balls The chances of fingerprint duplication, they preached, are one in a billion, and though there are more than two-and-a-half billion persons on this earth, no one person has ever been found to duplicate another person in all physical characteristics.”).

⁵⁶ See Mnookin, *supra* n. 13, 75 Brook. L. Rev. at 1265-66, 1275 (arguing that exclusion “should have a central place in the judicial toolkit,” that judges including those evaluating the ACE-V method have been “lulled by plausible descriptions and

Criminal cases, in which personal liberty is at stake, ought to warrant at least as stringent an application of *Daubert* and *Kumho* as in civil cases.⁵⁷ When *Daubert*, *Kumho*, or even *Frye* is stringently—and

seemingly persuasive explanations of forensic science techniques into dismissing the importance of the nearly complete lack of empirical support for the experts' claims," and that advancements in the research base underlying courtroom forensics will depend on "in significant part on judges, and whether they are prepared, at long last, to evaluate pattern identification evidence with their eyes wide open and their heads out of the sand.")

⁵⁷ See Gertner, *supra* n. 34, at 10 ("We, the courts, can do better. In fact, we already do, albeit in civil, not criminal, cases

appropriately—applied, there should be no place in the courtroom for fingerprint examiners to make individualization claims, at least until the fingerprint examiner community and a community of relevant scientists are in consensus that research exists to back up the validity of an opinion that will be so compelling to jurors who have been told for their entire lives that fingerprint identification is infallible.

. . . . It makes no sense to ignore *Daubert* and *Kumho* when liberty is at stake, but apply those cases rigorously when all that is involved is money. We must do better.").

How I Created a Monster

GABRIEL A. FUENTES

The author is a partner with Jenner & Block LLP, Chicago.

I never thought that as a trial lawyer, I would ever have anything in common with the creator of the monster in Mary Shelley's *Frankenstein*. But like Shelley's "modern Prometheus," I learned that a natural, human striving for a greater understanding of scientific knowledge can have unintended and even dangerous consequences.

In my case, a 2009 pro bono court appointment in a federal drug prosecution, I dared to challenge a gold standard of so-called "scientific" evidence. I dared to argue that everything we have been told about fingerprint examiners' ability to claim a match between a fingerprint deposited on a surface somewhere and a particular individual is without any real scientific basis, and that without an adequate showing of validity or reliability these claims should not be admitted as proper expert testimony.

Immersed in the task and thrill of the exploration, I thought I was pioneering, and in many senses, I was. I thought I was on the cutting edge of science or law, or even both. Only after my creation came to life and began terrorizing the countryside did I realize the horror I had wrought.

Others had challenged fingerprint identification testimony before, but the courts widely rejected all of those challenges under a line of cases dating back at least 100 years. Those cases stretched to a time when the notion of the infallibility of fingerprint identification first entered the public consciousness

following the publication of Mark Twain's *Pudd'nhead Wilson* in 1894. In Twain's story, the protagonist lawyer wins the acquittal of his clients by matching the fingerprints on the murder weapon to someone else. Twain apparently had read the work of Sir Francis Galton, published two years earlier. Galton, in short, had dropped cut-up images of fingerprints onto a piece of paper and concluded that the chances of one person's print being identical to that of another specific person were 1 in 64 billion.

None of those earlier court challenges had the benefit of a watershed 2009 report by the National Academy of Sciences (NAS), the leading scientific research arm of the U.S. government. See NAT'L RESEARCH COUNCIL, NAT'L ACAD. OF SCIS., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (Feb. 2009), available at www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf. The NAS report concluded that, with the exception of nuclear DNA evidence, "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." *Id.* at 7. The report says, among other things, that fingerprint examiners are "unjustified" in claiming they can match a latent fingerprint to a particular individual, that broad claims about the uniqueness of individual prints across the population are unsubstantiated, and that the most commonly

accepted fingerprint identification method “is not specific enough to qualify as a validated method for this type of analysis” and does not guarantee that two analysts following the method will reach the same results. *Id.* at 142–44.

In my case, *United States v. Clacy Watson Herrera*, the purported match stemmed from two smudged fingerprints recovered from a piece of tape on a package of heroin U.S. Customs agents extracted from a courier’s body cavity in 1999 at a Texas airport. The indictment alleged that my client provided the courier with the heroin in Panama. What was more important, the government wanted to use fingerprint identification testimony to shore up a case based almost entirely on cooperating witnesses who implicated my client after pleading guilty. According to the government, the fingerprints showed not only that my client had handed the heroin package to the courier but also that everything said about him by other witnesses was true—that he was the source of supply in Panama for a conspiracy to smuggle liquefied narcotics into the United States in baby formula cans, emptied of formula and refilled with drugs.

The conspiracy, the government said, involved flying young women from the South Side of Chicago to Panama, where they would be handed the drug-filled baby formula cans. To make their transportation of the cans appear legitimate, the young women allegedly posed as the mothers of infants. The story got even worse: The alleged conspiracy actually *rented babies* from desperate South Side couples so couriers could bring real children back and forth to Panama as a cover for the scheme.

The charged plot was as dastardly as it was brilliant. The only question was whether the government could prove my client was behind it—without surveillance testimony, pictures, recorded interceptions, or any incriminating statements.

We would not hear about fingerprints until November 2009, on the day before the scheduled trial. I had asked the government to disclose the results of examinations and tests under Federal Rule of Criminal Procedure 16, and the government said there was none concerning fingerprints. Late in the pre-trial preparation, though, the government informed us that it was having some of the old evidence examined. I knew right away why—when I had been a prosecutor, I had done the same thing. If the examination results yielded nothing, that didn’t matter. What mattered was stopping defense counsel from telling the jury that the government never had the evidence examined for fingerprints at all. Juries have this thing about fingerprints. They believe in them, so not bothering to look for them would be potentially disastrous for the government.

Before too long, the government’s examiner reported back that, in his opinion, two latent prints on the drug packaging tape belonged to my client. I geared up to challenge the report, and that was when the monster was conceived.

Illustration by Tim Foley

Challenging Fingerprints

I reread the NAS report and its conclusion about the shortcomings of the government's method of fingerprint identification and of the supporting research in general. Building on its conclusion that more research was needed to support claims by examiners that they could "match" a latent fingerprint to a single person, I embarked on what may have been the first comprehensive *Daubert* challenge to the admissibility of fingerprint identification evidence based on the NAS report.

By the time it was filed, the *Daubert* motion and exhibits ran to hundreds of pages. I relied heavily on the NAS report to show that the hundred-year-old line of cases was simply wrong. No one had ever offered research validating fingerprint matching, and no court had ever really demanded any. Now the NAS had established that little or none existed. Some courts had relied on now-disproved testimony that fingerprint identification had a zero error rate. But none had undertaken a *Daubert* analysis of the issue since the NAS report except in one case in federal district court where I thought the issue had not been comprehensively presented. Therefore, the older cases were more than wrong; they were obsolete and failed to account for the evolving state of scientific knowledge. Relying on them would be a failure to acknowledge that science itself progresses.

Today's courts, I argued, should look at fingerprint evidence through a new prism. It wasn't that fingerprint identification, as a science, is unreliable. The problem, I sought to explain, is that prosecutors cannot make *Daubert's* required showing that the evidence has *already been shown* to be reliable. It was impossible to know whether the opinion in our case was reliable enough for the jury to hear. There was no real way to know whether the opinion of a match was well grounded or whether it was one of a growing list of misidentifications, including the widely reported case of Brandon Mayfield, an Oregon lawyer erroneously identified as a terrorist whose fingerprint was on a bomb fragment from the 2004 Madrid train bombing. A 2006 report by the Federal Bureau of Investigation's Office of Inspector General found a number of the fingerprint identification procedures in the Mayfield case lacking, and it recommended a number of reforms. OFFICE OF INSPECTOR GEN., FED. BUREAU OF INVESTIGATION, A REVIEW OF THE FBI'S HANDLING OF THE BRANDON MAYFIELD CASE (Mar. 2006), available at www.justice.gov/oig/special/s0601/PDF_list.htm. Almost none of those reforms had been implemented by the Department of Homeland Security lab that had done the testing in my case.

The government responded that the NAS report changed nothing and that courts had admitted fingerprint identification for 100 years. Moreover, the government said, fingerprint identification wasn't a science anyway. Rather, it's an "experiential" discipline, meaning that the government need not provide

any research at all to support its scientific validity because the evidence did not need scientific endorsement to be admissible. And no such research was presented. Sir Francis Galton was not even mentioned. My expert affidavit, from a researcher whose work was cited in the NAS report, went un rebutted.

The district court denied the *Daubert* motion, following the pre-NAS report case law. We proceeded to trial, at which the fingerprint evidence, not surprisingly, took center stage. After the defense lost a battle royal over whether the packaging tape should be admitted, the government's fingerprint identification expert took the stand to give what the government had earlier called nonscientific testimony about my client being the person who left the fingerprints on the drug packaging.

In his first few words to the jury, the examiner identified himself as "a forensic scientist" assigned to the Department of Homeland Security's "Southwest Regional Science Center" in Houston. So much for the discipline being "experiential."

Different examiners looking at the same sets of fingerprints reached different conclusions.

He blew one of the latent prints up onto a large TV screen and launched into a whirlwind narrative about why it was a "compelling" match to the defendant. The jury seemed to get it—here was somebody who knew a lot more than they did about fingerprints, and he was saying the prints came from the defendant.

The cross-examination, I knew, represented another step on our journey toward making a record I hoped some reviewing court would ultimately see was devoid of anything demonstrating the reliability or validity of this evidence. The district court allowed me great leeway in presenting not only the NAS report and the Brandon Mayfield episode but also the expert cited by the NAS for the proposition that no study established the validity or reliability of the government's method.

Our expert, Ralph Haber, testified that he chose to examine the research basis behind fingerprints because he "figured there would be a hundred years of research and data that we could study." Instead, he said, "we were unable to find a single experiment, a single article, a single book that had published any research . . . on how accurate those methods were or how accurate an individual fingerprint examiner was." But there was

research, and Haber cited it, showing that different examiners looking at the same sets of fingerprints reached different conclusions, while other research showed that the same examiners looking at the same evidence at different times also reached different conclusions. Haber mentioned another study, by a researcher named Itiel Dror, showing that once “biasing information” suggesting that the identified person could not have left the fingerprint was told to examiners, 80 percent of them changed their answer, five years later, from identifications to exclusions.

Haber also testified that a proper evaluation of an examiner’s identification testimony required analysis of the examiner’s methods, and such an analysis could not be done if the examiner had documented only the conclusions. A few days earlier, the government’s examiner had testified that he did not write down how he had applied his examination method because “[t]hat’s just not something that our laboratory requires us to do.”

My creature was walking and talking now, even if, as a legal concept, it didn’t sway jurors who had probably been bombarded their entire lives with fingerprint identification’s established place in popular culture. After all, hadn’t the Supreme Court said in *Daubert* that the reliability of expert testimony is for the court, not the jury? In the end, after a bruising four-week trial, the jury convicted my client on eight of the 14 counts, including the one involving the heroin package with the latent fingerprints said to be left by him and no one else.

The Appeal

Despite the verdict, my creature was just getting started. We proceeded to the Seventh Circuit. The fingerprint match to our client became the central issue in the appellate briefing and the oral argument. I hoped the long-standing authority accepting fingerprint identification testimony more or less unquestioningly was ripe for a fresh look after the NAS report. *See* Gabriel A. Fuentes, *Toward a More Critical Application of Daubert in Criminal Cases: Fingerprint Opinion Testimony After the National Academy of Sciences Report*, BLOOMBERG BNA EXPERT EVIDENCE REP. (Oct. 22, 2012), available at www.bna.com/uploadedFiles/Content/Press/Toward_a_More_Critical_Application_of_Daubert_EXER.pdf (summarizing evolution of case law on admissibility of fingerprint opinion testimony before Seventh Circuit’s decision in *United States v. Herrera*, 704 F.3d 480 (7th Cir.), *cert. denied*, 134 S. Ct. 175 (2013)).

Alas, my creature didn’t fare well on appeal. The Seventh Circuit’s opinion, written by Judge Posner, reached several key conclusions:

- “Responsible” fingerprint matching testimony by an expert “is admissible evidence, in general and in this case.” While the opinion did not make clear precisely what could render

such testimony “responsible,” it suggested that certification by an examiners’ group called the International Association for Identification might be one factor. *Herrera*, 704 F.3d at 486–87.

- Fingerprint matching is not as reliable as DNA matching, but it need not be because other forms of admissible identification evidence, such as eyewitness testimony, is “less rigorous than the kind of scientific matching involved in DNA evidence,” and, moreover, “[e]vidence doesn’t have to be infallible to be probative.” *Id.* at 486.
- Comparing latent fingerprints to known fingerprints is analogous to an art expert opining on the style of a particular artist to conclude that a work of art is genuine or a forgery. *Id.* at 485–86.
- “[E]rrors in fingerprint matching by expert examiners appear to be extremely rare,” and the likelihood of error based on the probability of two people having “identical” fingerprints is low. *Id.* at 487. For this conclusion, the court relied heavily on its impression that Francis Galton in 1892 had estimated that the odds of “two people in the world having identical fingerprints” were 1 in 64 billion. *Id.* (Twenty-four scientists and scholars who later filed an amicus brief supporting Herrera’s petition for certiorari differed with this interpretation of Galton’s estimate, noting that it applied to whether one person’s fingerprint was identical to that of any other random person’s; the amici said Galton’s actual estimate of the probability that a specific fingerprint would match any other in the world was 1 in 4).

Naturally, I respectfully disagreed with each of these views and with the idea that a *de facto* *Daubert* hearing could be conducted on appeal without either party having its say about the validity of the court’s independent research. Although the issue was one of first impression, and the court of appeals did its usual conscientious and thorough job, I feared my creature had gotten out of hand. In the Mary Shelley novel, the monster eventually kills Victor Frankenstein’s wife and taunts him as he lies over the corpse. As I read the opinion and contemplated the death of my client’s claims, I felt a little like Victor Frankenstein.

The Questions Continue

The Supreme Court denied certiorari, leaving me to reflect on my creature and the unintended consequences of having created it. On the one hand, the monster left me far behind and is making mischief in the wider world, spreading to other areas of evidence law. The Seventh Circuit has now applied *Herrera* outside the fingerprint context to hold that an expert in premises security could testify even though his methodology “may not have been scientific.” *Lees v. Carthage Coll.*, 714 F.3d 516,

524–25 (7th Cir. 2013). But as the *Herrera* court noted correctly, even “non-scientific” expert evidence must be “shown to be reliable” before it may be admitted. *Herrera*, 704 F.3d at 486. *Herrera* left unanswered questions about how the reliability of “non-scientific” evidence will be judged, such as (1) when is an expert’s testimony sufficiently “non-scientific” to justify a less rigorous application of *Daubert*? and (2) when we say we need not be as rigorous, just how far will we drift from *Daubert*’s proven mechanism of ensuring reliability? *Daubert* was meant to be a flexible standard, but the heart of the case was about testing the validity of testimony from experts otherwise all too likely to be taken at face value by lay jurors.

My monster is also hard at work in trial courts around the country. Despite the NAS’s carefully researched and reasoned questioning of the legendary notion that a fingerprint examiner can confirm a match, prosecutors continue to offer this testimony in criminal cases and federal and state courts continue to admit it, sometimes relying on *Herrera*. A brief survey of state court decisions finds examples in Arizona, Illinois, Pennsylvania, and Washington. The Illinois case was particularly significant. The state court cited *Herrera*, without extensive discussion, for the reliability and admissibility of fingerprint identification evidence to uphold the conviction of one of the accused murderers of seven people in a Brown’s Chicken restaurant in January 1993. *People v. Luna*, 989 N.E.2d 655, 671 (Ill. App. Ct. 2013).

On the other hand, there’s hope my creature may yet find the right path. *Herrera* might spark a rethinking of the unquestioned admission of fingerprint matching or at least might prompt a reasoned discussion about the limits of the science—or, shall we say, the “experiential discipline.” Much has happened since the 2010 *Herrera* trial to provide grist for a decent *Daubert* or *Frye* hearing. Sources considered authoritative in the fingerprint examiner community have issued studies and position papers urging examiners to stop short of claims that they may “individualize” a latent fingerprint to a single person, in view of the inadequate research supporting that conclusion. See NAT’L INST. OF STANDARDS & TECH., NAT’L INST. OF JUSTICE, LATENT PRINT EXAMINATION AND HUMAN FACTORS: IMPROVING THE PRACTICE THROUGH A SYSTEMS APPROACH 197 (Feb. 2012), available at www.nist.gov/oles/upload/latent.pdf; Scientific Working Grp. on Friction Ridge Analysis, Study & Tech., Individualization/Identification Position Statement (Apr. 21, 2012), www.swgfast.org/Comments-Positions/120306_Individualization-Identification.pdf. Meanwhile, the value and methodology of more recent research on fingerprint misidentification rates are the subject of a robust debate. Compare Bradford T. Ulery et al., *Accuracy and Reliability of Forensic Latent Fingerprint Decisions*, 108 PROC. OF NAT’L ACAD. OF SCIS. 7733 (2011), available at www.pnas.org/content/108/19/7733.full, with R.N. Haber & L. Haber, *Experimental Results of Fingerprint Comparison*

Validity and Reliability: A Review and Critical Analysis, SCI. & JUST. (2014), available at [www.scienceandjusticejournal.com/article/S1355-0306\(13\)00085-3/abstract](http://www.scienceandjusticejournal.com/article/S1355-0306(13)00085-3/abstract).

To date, no court has sought to unpack, analyze, or revisit the rationale behind *Herrera*’s conclusions or the authoritative-ness of its independent research. That may also change. Some scholars already have begun to chisel away at *Herrera*’s application of *Daubert*. See Jonathan J. Koehler, *Forensic Fallacies and a Famous Judge*, 54 JURIMETRICS J. 216–17 (Spring 2014) (criticizing *Herrera*’s basis for concluding that fingerprint identification errors are “very rare”); DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 33:20 (Dec. 2013) (calling the evidentiary basis for *Herrera*’s determination of fingerprint identification’s reliability, made without a *Daubert* hearing, “remarkably thin,” and describing *Herrera*’s analogies between fingerprint identification and the detection of art forgery and the admissibility of eyewitness identification as “a strikingly peculiar passage in a case involving, let us remember, expert evidence”).

Unlike Victor Frankenstein, litigators can’t afford regret.

Other creative defense counsel should and probably will try the argument, and perhaps a searching *Daubert* inquiry will help a trial court sort out whether or not there really is a valid basis for expert claims of a fingerprint match or whether the testimony should be limited to the examiner’s description of similarities between latent and known fingerprint images. Perhaps someday a court will tell us more about the degree of rigor needed to assess “experiential” experts whose testimony, as the Seventh Circuit in *Herrera* correctly stated, still must be “shown to be reliable.”

In the end, I remind myself that the goal was never to achieve some lofty ideal or strike a blow for “good science.” I was simply defending a client. If the circumstances demanded it, I would do it again. Unlike Victor Frankenstein, litigators can’t afford regret. ■

2013 WL 3773550 (U.S.) (Appellate Petition, Motion and Filing)
Supreme Court of the United States.

Clacy Watson HERRERA, Petitioner,

v.

UNITED STATES OF AMERICA, Respondent.

No. 12-1461.

July 17, 2013.

On Petition for a Writ of Certiorari to the United States Court of Appeals for the Seventh Circuit

Brief of Amici Curiae Interested Scientists and Scholars in Support of Petition for Writ of Certiorari

Courtney Gilligan Saleski, Counsel of Record, Stanley J. Panikowski, Benjamin M. Gipson, DLA Piper LLP (US), One Liberty Place, 1650 Market Street, Suite 4900, Philadelphia, PA 19103, (215) 656-2431, courtney.saleski @dlapiper.com, Counsel for Amici Curiae, Interested Scientists and, Scholars.

***i TABLE OF CONTENTS**

TABLE OF CONTENTS	i
TABLE OF AUTHORITIES	iii
INTEREST OF <i>AMICI CURIAE</i>	1
SUMMARY OF ARGUMENT	5
ARGUMENT	7
I. THE SEVENTH CIRCUIT IMPROPERLY AUTHORIZED DISTRICT COURTS TO ABDICATE THEIR GATEKEEPING DUTIES IN DECIDING WHETHER TO ADMIT EXPERT TESTIMONY	7
II. THE SEVENTH CIRCUIT MISUSED <i>STARE DECISIS</i> TO STIFLE CURRENT SCIENTIFIC UNDERSTANDING ABOUT FINGERPRINT IDENTIFICATION ANALYSIS	10
III. THE SEVENTH CIRCUIT'S VIEW THAT FINGERPRINT EVIDENCE IS <i>PER SE</i> RELIABLE IS BASED ON UNSOUND REASONING AND OUTDATED PREMISES	13
A. Professional Certification Does Not Equal Reliability	14
B. Retrospective Analysis of Post-Conviction DNA Exonerations Does Not Provide Evidence of the Reliability of Fingerprint Analysis	15
C. The Seventh Circuit Misinterpreted Galton's Probability Analysis	16
*ii D. The "Appearance" that Fingerprint Errors Are Rare Does Not Constitute Evidence of the Reliability of Fingerprint Analysis	19
E. The Training of Fingerprint Analysts to Determine a Threshold at which Fingerprint Analysis Becomes Reliable Does Not Mean Reliability Has Been Achieved	20
CONCLUSION	20

***iii TABLE OF AUTHORITIES**

CASES

<i>Daubert v. Merrell Dow Pharmaceuticals</i> , 509 U.S. 579 (1993)	<i>passim</i>
<i>Grice v. State</i> , 151 S.W.2d 211 (Tex. Crim. App. 1941)	19
<i>Kumho Tire v. Carmichael</i> , 526 U.S. 137 (1999)	<i>passim</i>
<i>Melendez-Diaz v. Massachusetts</i> , 557 U.S. 305 (2009)	9
<i>State v. Kuhl</i> , 175 P. 190 (Nev. 1918)	19
<i>United States v. Havvard</i> , 117 F. Supp. 2d. 848 (S.D. Ind. 2000)	11
<i>United States v. Havvard</i> , 260 F.3d. 597 (7th Cir. 2001)	10, 11, 12
<i>United States v. Oliveiran</i> , No. 1:08-cr-10104-NG, D.I. 45, Procedural Order: Trace Evidence (D. Mass. Mar. 10, 2010)	9

OTHER AUTHORITIES

Bradford T. Ulery et al., <i>Accuracy and Reliability of Forensic Latent Fingerprint Decisions</i> , 108 Proceedings of the National Academy of Sciences 7733 (2011)	14
Brandon L. Garrett, <i>Convicting the Innocent: Where Criminal Prosecutions Go Wrong</i> (2011)	15
Brandon L. Garrett and Peter Neufeld, <i>Invalid Forensic Science Testimony and Wrongful Convictions</i> , 95 Virginia L. Rev. 1 (2009)	15, 16
*iv David A. Stoney and John I. Thornton, <i>A Critical Analysis of Quantitative Fingerprint Individuality Models</i> , 31 J. of Forensic Sciences 1187 (1986)	17
David L. Faigman et al., <i>Modern Scientific Evidence: Forensics</i> (2008)	11, 12
Elizabeth F. Loftus and Simon A. Cole, <i>Contaminated Evidence</i> , 304 Science 959 (2004)	15
Francis Galton, <i>Finger Prints</i> (1892)	17
Greg Hampikian, Emily West, and Olga Akselrod, <i>The Genetics of Innocence: Analysis of 194 U.S. DNA Exonerations</i> , 12 Annual Review of Human Genetics 97 (2011)	15
Jennifer L. Mnookin, <i>A Blow to the Credibility of Fingerprint Evidence</i> , Boston Globe (Feb. 2, 2004)	15
Mark Denbeaux and D. Michael Risinger, <i>Kumho Tire and Expert Reliability: How the Question You Ask Gives the Answer You Get</i> , 34 Seton Hall L. Rev. 15 (2003)	7
Michael J. Saks, <i>Reliability Standards - Too High, Too Low, or Just Right? The Legal and Scientific Evaluation of Forensic Science (Especially Fingerprint Expert Testimony)</i> , 33 Seton Hall L. Rev. 1167 (2003)	7
*v National Institute of Standards and Technology, <i>Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach</i> , Expert Working Group on Human Factors in Latent Print Analysis, U.S. Department of Commerce (2012)	8
National Research Council of the National Academy of Sciences, <i>Strengthening Forensic Science in the United States: A Path Forward</i> (2009) .	<i>passim</i>
Paul Giannelli, <i>Fingerprints Challenged!</i> , 17 Crim. Just. 33 (2002)	12
Simon A. Cole, <i>More Than Zero: Accounting for Error in Latent Fingerprint Identification</i> , 95 J. Crim. L. & Criminology 985 (2005)	15, 16
Simon A. Cole, <i>Suspect Identities: A History of Fingerprinting and Criminal Identification</i> (2001)	18
Stephen M. Stigler, <i>Galton and Identification by Fingerprints</i> , 140 Genetics 857 (1995)	17

*1 INTEREST OF AMICI CURIAE¹

Pursuant to Supreme Court Rule 37.3, the interested scientists and scholars listed below, as *amici* *2 *curiae*, respectfully submit this brief in support of the petition for writ of *certiorari*. The interested scientists and scholars ask this Court to hold that fingerprint identification evidence must meet the same standard of reliability that governs all expert testimony under *Kumho Tire v. Carmichael*, 526 U.S. 137 (1999). The National Research Council's 2009 report, *Strengthening Forensic Science in the United States: A Path Forward*, called into question the reliability of this type of evidence. This report underscores the need to rigorously examine the reliability of fingerprint identification evidence and reject the Seventh Circuit's *per se* rule of admissibility.

Amici are an international group of 24 scientists, scholars, and attorneys who are familiar with evidence law and the scientific literature concerning forensic science and fingerprint identification. *Amici* believe that the Seventh Circuit wrongly decided important questions of law in this case. *Amici* write to inform the Court of the need to review the Seventh Circuit's decision.

Amici consist of the following:

- David Balding is Professor of Statistical Genetics at University College London. From 2008 to 2012 he was a member of the External Advisory Group of the UK Forensic Science Service.
- Robert Bradley is Professor Emeritus of Politics and Government at Illinois State University.
- Simon A. Cole is Professor of Criminology, Law & Society and Director of the Newkirk Center for Science and Society at the University of California, Irvine.
- *3 • Emma Cunliffe is Associate Professor of Law at the University of British Columbia in Canada and a regular instructor in expert evidence at the National Judicial Institute Evidence Workshop.
- Gary Edmond is Professor of Law and Director of the Program in Expertise, Evidence and Law, Network for Interdisciplinary Studies of Law at The University of New South Wales, Australia.
- Jules Epstein is Associate Professor of Law at Widener University School of Law (Delaware). He was a member of the NIST Expert Working Group on Human Factors in Latent Print Analysis.
- Jessica Gabel is an Associate Professor of Law at Georgia State University College of Law.
- Brandon Garrett is Roy L. and Rosamund Woodruff Morgan Professor of Law at the University of Virginia.
- Nancy Gertner is Professor of Practice, Harvard Law School and Judge, United States District Court for the District of Massachusetts (Ret.).
- Greg Hampikian is a professor of Biology, and Criminal Justice at Boise State University and Director of the Idaho Innocence Project.
- Richard I. Kemp is Associate Professor of Psychology at the University of New South Wales, Australia. He is a member of the Australian Academy of Forensic Sciences.
- Roger Koppl is Professor of Finance, Syracuse University and Faculty Fellow, Forensic and *4 National Security Sciences Institute, Syracuse University.
- Kristy Martire is a Senior Lecturer in the School of Psychology at the University of New South Wales.
- Joëlle Anne Moreno is Professor of Law at Florida International University.
- Geoffrey Stewart Morrison is Director of the Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications, University of New South Wales.
- Myrna Raeder is Professor of Law at Southwestern Law School.
- D. Michael Risinger is John J. Gibbons Professor of Law and Associate Director, Last Resort Exoneration Project, Seton Hall University School of Law.
- Andrew Roberts is Senior Lecturer in Law at the University of Melbourne, Australia.
- Michael J. Saks is Regents Professor at the Sandra Day O'Connor College of Law, Arizona State University.

- Mehera San Roque is a Senior Lecturer in Law and Director of the JD Program in the School of Law at the University of New South Wales.
- David Siegel is Professor of Law and Director of the Center for Law and Social Responsibility at New England Law | Boston.
- William C. Thompson is Professor of Criminology, Law & Society at the University of California, Irvine.
- *5 • John Vokey is Professor of Psychology at the University of Lethbridge, Canada.
- Joëlle Vuille is a Visiting Scholar in the Department of Criminology, Law & Society at the University of California, Irvine. She holds a PhD in criminology from the University of Lausanne, Switzerland.

SUMMARY OF ARGUMENT

The fact that courts have customarily admitted certain scientific evidence does not by itself make the evidence sufficiently reliable for use at a federal criminal trial. Scientific knowledge evolves. And sometimes the evolution of scientific understanding raises questions about a long-standing approach to a certain category of scientific evidence. Litigants should have an opportunity to challenge these older views in light of more recent studies. But here, the United States Court of Appeals for the Seventh Circuit eschewed this approach by adopting what amounts to a *per se* rule of admissibility for fingerprint evidence.

The Seventh Circuit's rule ossifies outdated scientific views as legal precedent and renders legal analysis of the reliability of fingerprint evidence virtually impervious to evolving scientific knowledge. This decision is inconsistent with the approach to the admissibility of expert evidence that this Court mandated in *Daubert* and *Kumho Tire*. The Seventh Circuit's decision not only affects a widely-used category of expert evidence, but also risks infecting other subjects of expert testimony. *Amici* therefore ask this Court to grant review on the important question of law concerning the framework for evaluating the reliability of fingerprint evidence that the petition for *certiorari* presents.

*6 This Court's decision in *Kumho Tire* requires a district court to evaluate proposed expert evidence to determine if it is sufficiently reliable for the task at hand. The proponent of the evidence bears the burden of showing that the evidence is reliable and relevant. In this case, the district court failed in its gatekeeping duties and allowed fingerprint evidence to be introduced, not after evaluation under *Kumho Tire*, but instead because such evidence has been customarily admitted in the past. In doing so, the district court did not properly account for a 2009 report by the National Research Council, the arm of the National Academies of Science tasked with reporting on scientific matters, that called into the question the reliability of fingerprint evidence. That report concluded that the "ACE-V" method used by examiners - the method used to allegedly match Petitioner's fingerprints here - was not validated; that only limited information could be found in scientific literature about the reliability of fingerprint analysis; and that the type of fingerprint testimony offered in this case ("individualization") was not supported.

The Seventh Circuit erred as a matter of law when it endorsed the district court's failure to properly discharge its gatekeeping duties. The Seventh Circuit also overstepped its role as an appellate court by making its own determination of the reliability of fingerprint evidence - a determination that, in turn, was itself based on flawed premises and logic. This Court should grant review and reversal of the Seventh Circuit's judgment because it conflicts with this Court's precedents on a question of widespread importance to the law of evidence and conduct of criminal trials.

*7 ARGUMENT

I. THE SEVENTH CIRCUIT IMPROPERLY AUTHORIZED DISTRICT COURTS TO ABDICATE THEIR GATEKEEPING DUTIES IN DECIDING WHETHER TO ADMIT EXPERT TESTIMONY

The trial court is the gatekeeper for expert evidence.² *Kumho Tire v. Carmichael*³ requires the proponent of expert evidence to establish the relevance and reliability⁴ of that evidence for the “task at hand.”⁵ The trial court must ensure that the proponent satisfies its burden.⁶

The Seventh Circuit's decision in this case cannot be squared with these well-accepted principles. Here, *8 Petitioner asked that the government meet its burden of showing the reliability of fingerprint analysis under the circumstances of the case. In response, the government did not offer any evidence, studies, or data demonstrating the reliability of fingerprint analysis for the task at hand. Instead, the government's showing consisted of the citing of dated cases finding fingerprint evidence admissible.

Petitioner, by contrast, introduced recent evidence from the scientific community. This evidence included a 2009 report on forensic science by the National Research Council (NRC), the research arm of the National Academy of Sciences (NAS).⁷ The NAS is often called the most prestigious scientific institution in the United States, with an official mandate from the United States Congress to report on scientific matters. Petitioner also offered a 2012 report jointly commissioned by the National Institute of Standards and Technology and the National Institute of Justice.⁸

These scientific reports are highly relevant to the question of the reliability of fingerprint analysis. Three conclusions in the NAS Report are especially relevant to this case. First, the NAS Report concluded that the “ACE-V” method used by fingerprint examiners was not validated.⁹ Second, this report was able to find only “limited” information about the *9 accuracy and reliability of fingerprint analysis.¹⁰ Third, this report found no support for the fingerprint testimony of “individualization” - the testimony proffered in this case.¹¹

This Court has previously recognized the value of the NAS Report on the issue of the reliability of forensic science.¹² Relying on this study, the Court has explained that “[f]orensic evidence is not uniquely immune from the risk of manipulation” and that with many types of forensic evidence, including latent fingerprint analysis, “[t]here is wide variability ... with regard to techniques, methodologies, reliability, types and numbers of potential errors, research, general acceptability, and published material.”¹³

Despite all this, neither the district court nor the Seventh Circuit engaged in the rigorous analysis of reliability required by *Kumho Tire*. Denying Petitioner's request for a *Daubert* hearing,¹⁴ the district court did not permit a full inquiry into the reliability of fingerprint analysis for the task at hand.¹⁵ In doing so, the district court did not rely on scientific evidence, studies, or data, but instead rested its opinion on earlier cases finding fingerprint evidence to be admissible. Citing the same case law, the Seventh *10 Circuit approved the district court's decision to admit the evidence without rigorously testing its reliability.

The Seventh Circuit's approach contradicts *Kumho Tire's* requirement to rigorously test the reliability of expert evidence before deciding its admissibility. Cases approving the admission of similar evidence in the past cannot substitute for an evaluation of the evidence in light of current scientific understanding under the particular circumstances of the case at hand. This Court has provided a mechanism - *Daubert* - to conduct precisely such an evaluation. The Seventh Circuit has now constructed a bypass around *Daubert*, its progeny, and the Federal Rules of Evidence animating those decisions. This Court should grant review to dismantle this bypass and ensure that all expert evidence must pass through the gates that this Court's precedents have established.

II. THE SEVENTH CIRCUIT MISUSED *STARE DECISIS* TO STIFLE CURRENT SCIENTIFIC UNDERSTANDING ABOUT FINGERPRINT IDENTIFICATION ANALYSIS

Both the district court and Seventh Circuit relied on the Seventh Circuit's 2001 opinion in *United States v. Havvard*.¹⁶ But *Havvard* was decided before new scientific research called into question the type of fingerprint analysis at issue in this case. The

lower courts' reliance on this opinion illustrates a fundamental flaw that pervades their decisions: the misuse of *stare decisis* to ossify outdated scientific understandings as legal precedent. And the solution to this problem is simple: reverse the Seventh Circuit's decision in this case and require the application *11 of *Kumho Tire* and *Daubert*. This Court's precedents mandating a dynamic approach to the reliability of scientific evidence are precisely the antidote to the erroneous, static approach of the lower courts in this case.

As the district court in this case noted, the *Havvard* court's finding that fingerprint analysis was reliable rested upon three arguments: (1) the "adversarial testing" argument; (2) the fact that fingerprint conclusions are "verified" by a second examiner; and (3) the *Havvard* court's conclusion, based solely on the state's expert's un rebutted characterization of the error rate as "zero,"¹⁷ that the error rate was "exceptionally low."

Regarding "adversarial testing," the *Havvard* ruling was singled out for withering criticism in Petitioner's principal exhibit, the NAS Report, for its reliance on this concept.¹⁸ In particular, the Report implicitly criticized the "statement that fingerprinting met the *Daubert* testing criterion by virtue of having been tested by the adversarial process over the last one-hundred years," first made by the *Havvard* court. The NAS Report specifically found it "telling" that the legal treatise *Modern Scientific Evidence* dismissed the *Havvard* court's statement as "silly."¹⁹ In *12 fact, most scientists would view the notion that the use of scientific evidence in a criminal proceeding provides a "test" of its reliability is a gross distortion of the notion of "scientific testing" described in *Kumho Tire* and *Daubert*.

Further, evidence scholars have sharply criticized the *Havvard* court's "adversarial testing" argument.²⁰ These scholars have observed that the adversarial process of the courtroom is not an adequate substitute for scientific validation research. The district court here was well aware of the NAS Report's criticism of the *Havvard* court's reasoning. But it still concluded that the NAS Report had not "undermined" the *Havvard* court's reasoning on adversarial testing. (Pet. App. 42a.) This conclusion was flawed.

Moreover, the district court acknowledged that the NAS Report had effectively rebutted *Havvard's* "error rate" argument by pointing out serious deficiencies in evidence regarding error rates in fingerprint analysis. This concession is especially significant because one could reasonably say that the error rate of fingerprint identification *is* the reliability of fingerprint identification. The reliability of fingerprint evidence depends critically on the error rate of fingerprint examiners; without considering the known or potential rate of error, a court cannot assess the reliability of the evidence. Aware of this problem, the district court nonetheless admitted the fingerprint evidence, and the Seventh Circuit approved it. The lower courts instead should have applied this Court's precedents requiring proper scrutiny of scientific evidence before it is admitted.

*13 This Court should not let stand the untenable proposition that criminal proceedings serve as scientific tests of the reliability of the evidence used in those proceedings, foreclosing the possibility of a *Daubert* inquiry before introducing this evidence through expert testimony at trial. Nor should this Court allow the lower courts to misuse *stare decisis* to ignore later-arising scientific evidence that goes to the very heart of the reliability of expert testimony. Accordingly, review should be granted in this case.

III. THE SEVENTH CIRCUIT'S VIEW THAT FINGERPRINT EVIDENCE IS *PER SE* RELIABLE IS BASED ON UNSOUND REASONING AND OUTDATED PREMISES

Rather than hold the proponent of the fingerprint evidence to its burden of establishing reliability, the Seventh Circuit itself concluded that this type of expert evidence is reliable. The court's substitution of its own reliability determination when the proponent of expert evidence failed to meet its burden oversteps the role of an appellate court. The court then compounded the error by basing its determination on flawed logic and premises.

The Seventh Circuit's *sua sponte* determination that fingerprint identification evidence is reliable rested on five assertions: (1) the experts proffered by the government were certified by the International Association for Identification (IAI) (Pet. App. 13a);

(2) none of the first 194 prisoners exonerated by the post-conviction DNA testing in the United States was convicted on the basis of fingerprint evidence (Pet. App. 14a); (3) “[t]he probability of two people in the world having identical fingerprints” has been estimated by Francis Galton to be 1 in 64 billion (*id.*); *14 (4) “errors in [fingerprint] matching appear to be very rare” (*id.*); (5) “[e]xaminers' training includes instruction on how to determine whether a latent print contains enough detail to enable a reliable matching to another print” (*id.*).

None of these five points, either individually or in the aggregate, can support an inference about the reliability of expert testimony by a fingerprint examiner in this case. Moreover, some of these points are factually inaccurate. Most of all, none of these points can support the Seventh Circuit's conclusion that the government's fingerprint identification evidence could be admitted through expert testimony without a *Daubert* analysis.

A. Professional Certification Does Not Equal Reliability

The fact that an occupation runs a certification program does not constitute evidence about how accurately (or “reliably”) members of that occupation perform various tasks. The IAI Latent Print Certification program is a commendable endeavor, but its existence does not constitute evidence of the reliability of the government's examiners in performing the task undertaken in this case. Indeed, preliminary data on the reliability of latent print identification does not support the notion that certified examiners perform better than uncertified examiners.²¹ Accordingly, professional certifications have no bearing on whether the fingerprint analysis before the court is reliable evidence.

*15 B. Retrospective Analysis of Post-Conviction DNA Exonerations Does Not Provide Evidence of the Reliability of Fingerprint Analysis

The Seventh Circuit's assertion that “Of the first 194 prisoners in the United States exonerated by DNA evidence, none had been convicted on the basis of erroneous fingerprint matches” (Pet. App. 14a) is false. Stephan Cowans, the 141st person exonerated by post-conviction DNA testing in the United States, was convicted in large measure on the basis of erroneous fingerprint evidence.²²

Moreover, using post-conviction DNA exonerations to make inferences about the reliability of fingerprint identification is problematic because post-conviction *16 DNA exonerations are not a representative sample of cases in which fingerprint evidence is used.²³ The erroneous belief that the reliability of evidence can be inferred from its prevalence among post-conviction DNA exonerations is a common one. But it constructs a proportion entirely different from that which would be required to estimate an “error rate.” An error rate would be based on the number of errors divided by the number of attempts. Using post-conviction exonerations constructs a proportion with both a different denominator - post-conviction exonerations - and a different numerator - post-conviction exonerations of individuals who were “convicted on the basis of erroneous fingerprint matches.” This latter calculation does not yield the error rate for fingerprint analysis. The Seventh Circuit's views on this issue likewise do not support its *per se* rule of admissibility.

C. The Seventh Circuit Misinterpreted Galton's Probability Analysis

The Seventh Circuit's claim that “[t]he great statistician Francis Galton” estimated “[t]he probability of two people in the world having identical fingerprints” as 1 in 64 billion (Pet. App. 14a) is incorrect. This figure was the estimate Galton gave for the probability that one particular “fingerprint” (more correctly “friction ridge skin pattern”) is “identical” to *another particular fingerprint*.²⁴ Thus, for example, 1 *17 in 64 billion would have been Galton's estimated probability that the friction ridge skin of the left index finger of Petitioner Clacy Herrera is identical to the friction ridge skin of the left index finger of the latent print examiner who testified in this case.²⁵ Galton's estimate for the probability that a specified fingerprint would be identical to *any* other fingerprint in the world population (which Galton estimated at 1.6 billion in 1892) was a far more modest 1 in

4.²⁶ Thus, Galton's estimate of the probability that the fingerprint of the left index finger of Clacy Herrera would be identical to *some other unspecified fingerprint* in the (1892) world

The Seventh Circuit's error is a common one. It confuses the probability that two specified people possess the same attribute with the probability that an attribute possessed by one person is also possessed by at least one other *unspecified* person in a population of persons. The error is similar to the well-known “birthday problem” in statistics. The probability that two specified people share a birthday is much lower than the probability that one specified person shares a birthday with at least one unspecified member of a group of people.

“The probability of two people in the world having identical fingerprints” (Pet. App. 14a) is not the probability relevant to the issue the Seventh Circuit was being asked to resolve. The relevant probability is the probability of finding the features that were found in common between Petitioner's known prints and the latent prints adduced into evidence against him if someone other than Petitioner were the source of those latent prints. It is well understood in the literature, and it was stated in the NAS Report, that neither Galton's estimate nor any estimate of the probability of exact duplication addresses this question.²⁸

*19 There is a long tradition in U.S. case law of citing - and citing incorrectly²⁹ - Galton's and other researchers' estimates of the probability of exact duplication of fingerprints in support of the reliability of fingerprint analysis.³⁰ Granting *certiorari* and reversing the judgment will, among other things, stop the perpetuation of this error in the law.

D. The “Appearance” that Fingerprint Errors Are Rare Does Not Constitute Evidence of the Reliability of Fingerprint Analysis

The Seventh Circuit's statement that “errors in [fingerprint] matching appear to be very rare” (Pet. App. 14a) is nothing more than a vague *ipse dixit* assertion. Its opinion gives no detail on what evidence, studies, or data generated this “appearance” nor to what specific fingerprint “matching” task it is referring, nor of what sort of reliability rate is implied by the verbal formulation “very rare.” Indeed, NAS Report concluded that information on the error rate of fingerprint identification was lacking in 2009.³¹ Given the fact that the National Academy of Sciences could not find adequate data to claim that the error rate is very low for fingerprint *20 identification, it is unsurprising that the Seventh Circuit was unable to point to any authority for its *ipse dixit* assertion. Accordingly, this premise of the court's decision was based not on scientific opinion, but rather on the inertia of a system that has historically allowed this evidence to be presented.

E. The Training of Fingerprint Analysts to Determine a Threshold at which Fingerprint Analysis Becomes Reliable Does Not Mean Reliability Has Been Achieved

The mere fact that examiners are trained to try to determine when the quality of a latent print reaches some threshold level at which fingerprint analysis becomes “reliable” (Pet. App. 14a), does not mean that the training was successful. Further information about the actual reliability of fingerprint analysts at performing the task at hand would be required. That type of information would be one of the subjects of a *Daubert* inquiry. But in the absence of any such inquiry here, the Seventh Circuit improperly relied on training as a proxy for reliability. This is yet another reason why the decision below is flawed and why this Court's intervention is required to ensure its precedents are given full effect in this area of law.

CONCLUSION

“[T]rial-court discretion in choosing the manner of testing expert reliability ... is not discretion to abandon the gatekeeping function.”³² *Amici* believe it serves neither science nor law to admit fingerprint *21 identification evidence without a rigorous *Daubert* inquiry to determine its reliability under the particular facts of the case. The Seventh Circuit's decision improperly

replaces this Court's precedent with a *per se* rule of admissibility in this controversial area of forensic analysis. The Seventh Circuit's rule places a decisive and unvarying thumb on one side of the scale in assessing the admissibility of fingerprint identification evidence. Its decision also stakes out an approach that could improperly infect other areas of evidence law. *Amici* therefore urge the Court to grant *certiorari* to review and reverse the judgment.

Footnotes

- 1 Pursuant to Rule 37.6, *amici* certify that no counsel for a party authored this brief in whole or in part, and that no such counsel or party, other than *amici*, its members, or its counsel made a monetary contribution to the preparation or submission of this brief. Counsel of Record for all parties received timely notice of the intent to file this brief. Letters from the parties consenting to the filing of this brief have been filed with the Clerk of this Court.
- 2 Michael J. Saks, *Reliability Standards - Too High, Too Low, or Just Right? The Legal and Scientific Evaluation of Forensic Science (Especially Fingerprint Expert Testimony)*, 33 *Seton Hall L. Rev.* 1167, 1173-74 (2003) (“Elementary principles of law place the burden of proof on the proponent of the admission of evidence. Accordingly, *Daubert* places the initial burden of production on the proponent of the proffered expert witness....”)
- 3 *Kumho Tire v. Carmichael*, 526 U.S. 137 (1999).
- 4 This brief conforms to the usage in both the Seventh Circuit opinion and *Daubert v. Merrell Dow Pharmaceuticals* in which the term “reliability” is used to mean what many scientists would refer to as “validity.” As this Court noted in *Daubert*, “In a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*.” *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 590 n.9 (1993).
- 5 That is: the specific purpose for which the evidence is being offered. See Mark Denbeaux and D. Michael Risinger, *Kumho Tire and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 *Seton Hall L. Rev.* 15 (2003).
- 6 *Kumho Tire*, 526 U.S. at 147.
- 7 National Research Council of the National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press 2009) (“NAS Report”).
- 8 National Institute of Standards and Technology, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, Expert Working Group on Human Factors in Latent Print Analysis, U.S. Department of Commerce (2012).
- 9 NAS Report, *supra* note 7, at 142.
- 10 *Id.* at 142.
- 11 *Id.* at 7.
- 12 *Melendez-Diaz v. Massachusetts*, 557 U.S. 305, 318, 320-21 (2009).
- 13 *Id.*
- 14 *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993).
- 15 On the importance of such hearings for forensic science, see *United States v. Oliveiran*, No. 1:08-cr-10104-NG, D.I. 45, Procedural Order: Trace Evidence (D. Mass. Mar. 10, 2010).
- 16 *United States v. Havvard*, 260 F.3d. 597 (7th Cir. 2001).
- 17 *United States v. Havvard*, 117 F. Supp. 2d. 848, 854 (S.D. Ind. 2000).
- 18 NAS Report, *supra* note 7, at 96 (citing *Havvard* as an example of the proposition that “Federal 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”).
- 19 *Id.* at 103 (quoting *Modern Scientific Evidence*); see also David L. Faigman et al., *Modern Scientific Evidence: Forensics* (2008).
- 20 The treatise *Modern Scientific Evidence* was hardly alone in criticizing this reasoning. See, e.g., Paul Giannelli, *Fingerprints Challenged!*, 17 *Crim. Just.* 33, 35 (2002).
- 21 Bradford T. Ulery et al., *Accuracy and Reliability of Forensic Latent Fingerprint Decisions*, 108 *Proceedings of the National Academy of Sciences* 7733, 7735-36 (2011).
- 22 Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 *J. Crim. L. & Criminology* 985, 1014-16 (2005); Brandon L. Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*, 106-07 (2011); Brandon L. Garrett and Peter Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, *Virginia L. Rev.* 1, 73-74, 95 (2009); Elizabeth F. Loftus and Simon A. Cole, *Contaminated Evidence*, 304 *Science* 959 (2004); Jennifer L. Mnookin, *A Blow to the Credibility of Fingerprint Evidence*, *Boston Globe* (Feb. 2, 2004). It is true that the study cited by the Seventh Circuit reported no cases in which erroneous fingerprint evidence contributed to a wrongful conviction subsequently exposed by post-conviction DNA

testing. Greg Hampikian, Emily West, and Olga Akselrod, *The Genetics of Innocence: Analysis of 194 U.S. DNA Exonerations*, 12 Annual Review of Human Genetics 97 (2011). But the Seventh Circuit incorrectly characterizes this study as a study “[o]f the first 194 prisoners in the United States exonerated by DNA evidence” (Pet. App. 13). In fact, this was a study of a *sample* of 194 of the first 255 prisoners exonerated by post-conviction DNA testing. It appears that the Cowans case was incorrectly coded in this study. The first author of this study is a signatory to this brief.

- 23 See Brandon L. Garrett and Peter Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, Virginia L. Rev., 1, 21-22 (2009); Simon A. Cole (2005) *More Than Zero: Accounting for Error in Latent Fingerprint Identification* 95 J. Crim. L. & Criminology 985, 1025-28.
- 24 An estimate of the probability of the appearance of “identical” friction ridge skin patterns is meaningless without understanding the estimator’s rule for what degree and kind of similarity between two images will result in the images being deemed “identical.” Most such estimates are explicitly or implicitly based on rules in which the definition of “identity” is quite a bit less than, say, “indistinguishable in every detail down to the level of molecules or atoms.” That was certainly the case for Galton, for whom “identity” consisted merely of “minutiae” appearing in similar topographic locations in the two images, but it is true for all other estimates of the probability of identical friction ridge skin patterns as well. *18 population would have been 1 in 4.^[FN27] Galton’s estimate of the probability of finding a friction ridge skin pattern identical to Herrera’s in today’s larger world population would be even higher.
- 25 This probability might have been relevant if, for example, Herrera had hypothesized that some specified individual were the source of the print. But Herrera bore no burden to offer such a hypothesis, and he did not do so. In most cases, that specified individual would have fingerprints available for inspection, and the probability estimate would be moot. In cases, however, in which fingerprints from that specified individual were not available, the probability might be relevant.
- 26 Francis Galton, *Finger Prints*, 110-11 (1892); Stephen M. Stigler, *Galton and Identification by Fingerprints*, 140 Genetics 857, 858 (1995); David A. Stoney and John I. Thornton, *A Critical Analysis of Quantitative Fingerprint Individuality Models*, 31 J. of Forensic Sciences 1187, 1188 (1986); Simon A. Cole, *Suspect Identities: A History of Fingerprinting and Criminal Identification*, 80 (2001).
- 27 While it might be argued that the true potential donor pool of the latent prints on the drug packages in this case was less than the world population, which would imply a probability of duplication lower than 1 in 4, neither the government’s lawyers nor its expert witnesses defined such a reduced potential donor pool.
- 28 NAS Report, *supra* note 7, at 43, 144.
- 29 See, e.g., *Grice v. State*, 151 S.W.2d 211, 215 (Tex. Crim. App. 1941) (incorrectly reporting what was almost certainly Galton’s estimate as 1 in 64 million, rather than billion, and incorrectly reporting it as the population within which a duplication would be expected to occur, rather than as the probability of duplication between two specified fingers).
- 30 *Id.*; *State v. Kuhl*, 175 P. 190, 192 (Nev. 1918) (citing Balthazard’s claim that the probability of exact duplication is in 1 in 10⁶⁰).
- 31 NAS Report, *supra* note 7, at 142.
- 32 *Kumho Tire v. Carmichael*, 526 U.S. 137, 158-59 (1999) (Scalia, J., concurring).

OPINION



Bite-Mark Evidence Proving Unreliable

Despite a robust industry of experts in the field, criticism of the forensic technique is growing.

BY GABRIEL A. FUENTES

July turned out to be a bad month for bite-mark evidence. First, a high-ranking Obama administration official told a government-sponsored conference on managing forensic science error that bite-mark evidence should be “eradicated from forensic science.” Next, a Chicago federal judge issued perhaps the most stinging judicial rebuke yet of the unreliable and unsubstantiated notion that an “expert” can reliably associate a bite mark left in a crime victim’s skin with a particular person.

The latest criticisms came on the

heels of revelations in April by the FBI that 26 of 28 examiners in an elite FBI microscopic hair-comparison unit had overstated forensic matches in ways that favored prosecutors in 95 percent of trials in which they had offered evidence over two decades. It all served as a reminder of the need to rein in the runaway use of forensic “science” in our criminal courts.

Although the widely acclaimed 2009 report of the National Academy of Sciences highlighted the starkly limited reliability of many forensic disciplines commonly used to convict criminal defendants, many judges have been oddly

reluctant to acknowledge that the NAS report changed anything. But change may yet occur as the limits of certain forensic sciences receive wider acknowledgement.

In a speech to open the July conference on managing forensic error sponsored by the National Institute of Science and Technology, Jo Handelsman, associate director for science in the White House Office of Science and Technology Policy, cited studies showing that the conclusions of bite-mark examiners were “all over the place,” and said that such unreliable and unscientific methods “have to be eradicated from forensic science.”

On July 24, as the National Institute of Science and Technology conference in Arlington, Virginia, was wrapping up, U.S. District Judge Gary Feinerman released his memorandum opinion in *Starks v. City of Waukegan*, a civil rights action filed by an Illinois man who spent 20 years in prison for a rape before DNA exonerated him of that crime. Part of the earlier criminal trial evidence against Bennie Starks consisted of the opinion testimony of two dentists who told his jury that he was the source of bite marks left on the victim. Feinerman granted the dentists summary judgment in Starks' suit against them but did not mince words in finding bite-mark matching sorely lacking in reliability.

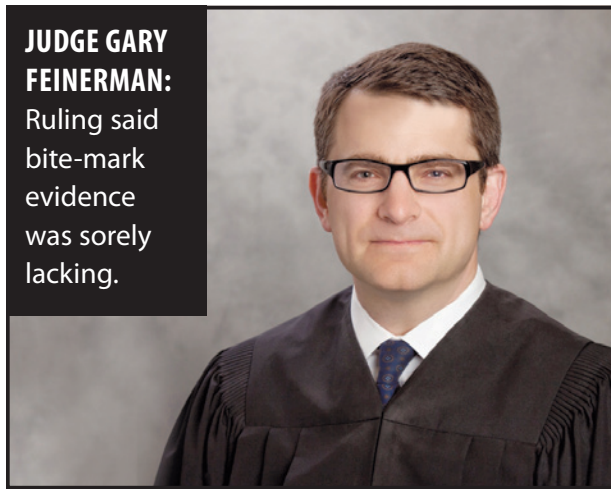
The judge began by citing the NAS report's conclusion that bite-mark identification was not supported by scientific studies, and that many of its underlying assumptions (such as the uniqueness of human dentition) have not been scientifically established. The evidence that does exist, Feinerman wrote, "is damning."

The judge specifically cited two peer-reviewed articles by prominent bite-mark identification critics whose research the NAS had relied upon for its report. Interestingly enough, those critics included Mary and Peter Bush, a married couple who led a group of researchers at the State University of New York at Buffalo, whose criticism of the orthodoxy of bite-mark identification triggered a backlash from professional bite-mark examiners and even a prosecutor.

The critics of bite-mark evidence included Dr. C. Michael Bowers, a California dentist who was the target of an unsuccessful campaign to expel him from the American Academy of Forensic Sciences based on trumped-up ethics charges. Feinerman's citation to the Bushes and Bowers represented not only vindication for them, but also a step forward, in that their

JUDGE GARY FEINERMAN:

Ruling said bite-mark evidence was sorely lacking.



research may have begun to gain traction in the courtroom.

As for whether bite-mark examiners themselves may authoritatively address the reliability of bite-mark evidence, Feinerman quoted Upton Sinclair: "It is difficult to get a man to understand something, when his salary depends on his not understanding it."

In concluding that the bite-mark examiners in the Starks case could not have deliberately falsified their testimony, the judge compared their opinions to those of astrologers or palm readers. It would not be deliberately false, the judge wrote, for an astrologer to say the defendant committed the crime because the planets were in a particular alignment, or for a palm reader to have

"grossly deviated from professional standards by mistaking the heart line for the head line."

Let's just say that the judge's opinion was not exactly a ringing endorsement of bite-mark identification evidence.

Feinerman's opinion, and the comments of Handelsman before the National Institute of Science and Technology, underscore the need for judges and practitioners to tread carefully when a self-proclaimed forensic "scientist" or "expert" comes to court with an opinion associating trace evidence with a criminal suspect. Is this evidence reliable? Do these experts, whether or not they claim their opinions to be grounded in "science" or mere "experience," have a reliable or valid basis for their opinions?

If mainstream scientists and judges are beginning to take note of the courtroom limits of weak forensic disciplines such as bite-mark comparison, perhaps there is hope that more courts will take the NAS report seriously and bring the needed rigor to their tasks as gatekeepers of expert evidence, upon which lay juries rely so heavily.

Gabriel A. Fuentes, a partner at Jenner & Block, has litigated forensic science matters including the defense of Dr. C. Michael Bowers before the American Academy of Forensic Sciences. The opinions here are his own.