

Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

The Federal Judicial Center contributed to this publication in furtherance of the Center's statutory mission to develop and conduct educational programs for judicial branch employees. The views expressed are those of the authors and not necessarily those of the Federal Judicial Center.

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

The development of the third edition of the *Reference Manual on Scientific Evidence* was supported by Contract No. B5727.R02 between the National Academy of Sciences and the Carnegie Corporation of New York and a grant from the Starr Foundation. The views expressed in this publication are those of the authors and do not necessarily reflect those of the National Academies or the organizations that provided support for the project.

International Standard Book Number-13: 978-0-309-21421-6

International Standard Book Number-10: 0-309-21421-1

Library of Congress Cataloging-in-Publication Data

Reference manual on scientific evidence. — 3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-309-21421-6 (pbk.)

ISBN-10: 0-309-21421-1 (pbk.)

1. Evidence, Expert—United States. I. Federal Judicial Center.

KF8961.R44 2011

347.73'67—dc23

2011031458

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

Copyright 2011 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Summary Table of Contents

A detailed Table of Contents appears at the front of each chapter.

- Introduction, 1
Stephen Breyer
- The Admissibility of Expert Testimony, 11
Margaret A. Berger
- How Science Works, 37
David Goodstein
- Reference Guide on Forensic Identification Expertise, 55
Paul C. Giannelli, Edward J. Imwinkelried, & Joseph L. Peterson
- Reference Guide on DNA Identification Evidence, 129
David H. Kaye & George Sensabaugh
- Reference Guide on Statistics, 211
David H. Kaye & David A. Freedman
- Reference Guide on Multiple Regression, 303
Daniel L. Rubinfeld
- Reference Guide on Survey Research, 359
Shari Seidman Diamond
- Reference Guide on Estimation of Economic Damages, 425
Mark A. Allen, Robert E. Hall, & Victoria A. Lazear
- Reference Guide on Exposure Science, 503
Joseph V. Rodricks
- Reference Guide on Epidemiology, 549
Michael D. Green, D. Michal Freedman, & Leon Gordis
- Reference Guide on Toxicology, 633
Bernard D. Goldstein & Mary Sue Henifin
- Reference Guide on Medical Testimony, 687
John B. Wong, Lawrence O. Gostin, & Oscar A. Cabrera
- Reference Guide on Neuroscience, 747
Henry T. Greely & Anthony D. Wagner
- Reference Guide on Mental Health Evidence, 813
Paul S. Appelbaum
- Reference Guide on Engineering, 897
Channing R. Robertson, John E. Moalli, & David L. Black
- Appendix A. Biographical Information of Committee and Staff, 961

Reference Guide on Forensic Identification Expertise

PAUL C. GIANNELLI, EDWARD J. IMWINKELRIED,
AND JOSEPH L. PETERSON

Paul C. Giannelli, L.L.M., is Albert J. Weatherhead III and Richard W. Weatherhead Professor of Law, and Distinguished University Professor, Case Western Reserve University.

Edward J. Imwinkelried, J.D., is Edward L. Barrett, Jr. Professor of Law and Director of Trial Advocacy, University of California, Davis.

Joseph L. Peterson, D.Crim., is Professor of Criminal Justice and Criminalistics, California State University, Los Angeles.

CONTENTS

- I. Introduction, 57
- II. Development of Forensic Identification Techniques, 58
- III. Reappraisal of Forensic Identification Expertise, 60
 - A. DNA Profiling and Empirical Testing, 60
 - B. *Daubert* and Empirical Testing, 62
- IV. National Research Council Report on Forensic Science, 64
 - A. Research, 66
 - B. Observer Effects, 67
 - C. Accreditation and Certification, 68
 - D. Proficiency Testing, 69
 - E. Standard Terminology, 70
 - F. Laboratory Reports, 70
- V. Specific Techniques, 71
 - A. Terminology, 71
- VI. Fingerprint Evidence, 72
 - A. The Technique, 73
 - B. The Empirical Record, 76
 - 1. Proficiency testing, 78
 - 2. The Mayfield case, 79
 - C. Case Law Development, 81
- VII. Handwriting Evidence, 83
 - A. The Technique, 83
 - B. The Empirical Record, 85
 - 1. Comparison of experts and laypersons, 86
 - 2. Proficiency studies comparing experts' performance to chance, 87

- C. Case Law Development, 89
- VIII. Firearms Identification Evidence, 91
 - A. The Technique, 91
 - 1. Firearms, 91
 - 2. Ammunition, 92
 - 3. Class characteristics, 92
 - 4. Subclass characteristics, 93
 - 5. Individual characteristics, 93
 - 6. Consecutive matching striae, 94
 - 7. Cartridge identification, 94
 - 8. Automated identification systems, 95
 - 9. Toolmarks, 96
 - B. The Empirical Record, 97
 - C. Case Law Development, 100
- IX. Bite Mark Evidence, 103
 - A. The Technique, 104
 - 1. Theory of uniqueness, 105
 - 2. Methods of comparison, 106
 - 3. ABFO Guidelines, 107
 - B. The Empirical Record, 108
 - 1. DNA exonerations, 109
 - C. Case Law Development, 110
 - 1. Specificity of opinion, 111
 - 2. Post-*Daubert* cases, 112
- X. Microscopic Hair Evidence, 112
 - A. The Technique, 112
 - B. The Empirical Record, 113
 - 1. Mitochondrial DNA, 116
 - 2. Proficiency testing, 116
 - 3. DNA exonerations, 117
 - C. Case Law Development, 117
- XI. Recurrent Problems, 120
 - A. Clarity of Testimony, 120
 - B. Limitations on Testimony, 121
 - C. Restriction of Final Argument, 124
- XII. Procedural Issues, 124
 - A. Pretrial Discovery, 125
 - 1. Testifying beyond the report, 126
 - B. Defense Experts, 127

I. Introduction

Forensic identification expertise encompasses fingerprint, handwriting, and firearms (“ballistics”), and toolmark comparisons, all of which are used by crime laboratories to associate or dissociate a suspect with a crime. Shoe and tire prints also fall within this large pattern evidence domain. These examinations consist of comparing a known exemplar with evidence collected at a crime scene or from a suspect. Bite mark analysis can be added to this category, although it developed within the field of forensic dentistry as an adjunct of dental identification and is not conducted by crime laboratories. In a broad sense, the category includes trace evidence such as the analysis of hairs, fibers, soil, glass, and wood. Some forensic disciplines attempt to individuate and thus attribute physical evidence to a particular source—a person, object, or location.¹ Other techniques are useful because they narrow possible sources to a discrete category based upon what are known as “class characteristics” (as opposed to “individual characteristics”). Moreover, some techniques are valuable because they eliminate possible sources.

Following this introduction, Part II of this guide sketches a brief history of the development of forensic expertise and crime laboratories. Part III discusses the impact of the advent of DNA analysis and the Supreme Court’s 1993 *Daubert* decision,² developments that prompted a reappraisal of the trustworthiness of testimony by forensic identification experts. Part IV focuses on the 2009 National Research Council (NRC) report on forensic science.³ Parts V through X examine specific identification techniques: (1) fingerprint analysis, (2) questioned document examination, (3) firearms and toolmark identification, (4) bite mark comparison, and (5) microscopic hair analysis. Part XI considers recurrent problems, including the clarity of expert testimony, limitations on its scope, and restrictions on closing arguments. Part XII addresses procedural issues—pretrial discovery and access to defense experts.

1. Some forensic scientists believe the word *individualization* is more accurate than *identification*. Paul L. Kirk, *The Ontogeny of Criminalistics*, 54 J. Crim. L., Criminology & Police Sci. 235, 236 (1963). The identification of a substance as heroin, for example, does not individuate, whereas a fingerprint identification does.

2. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993). *Daubert* is discussed in Margaret A. Berger, *The Admissibility of Expert Testimony*, in this manual.

3. National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (2009) [hereinafter NRC Forensic Science Report], available at http://www.nap.edu/catalog.php?record_id=12589.

II. Development of Forensic Identification Techniques

An understanding of the current issues requires some appreciation of the past. The first reported fingerprint case was decided in 1911.⁴ This case preceded the establishment of the first American crime laboratory, which was created in Los Angeles in 1923.⁵ The Federal Bureau of Investigation (FBI) laboratory came online in 1932. At its inception, the FBI laboratory staff included only firearms identification and fingerprint examination.⁶ Handwriting comparisons, trace evidence examinations, and serological testing of blood and semen were added later. When initially established, crime laboratories handled a modest number of cases. For example, in its first full year of operation, the FBI laboratory processed fewer than 1000 cases.⁷

Several sensational cases in these formative years highlighted the value of forensic identification evidence. The Sacco and Vanzetti trial in 1921 was one of the earliest cases to rely on firearms identification evidence.⁸ In 1935, the extensive use of handwriting comparison testimony⁹ and wood evidence¹⁰ at the Lindbergh kidnapping trial raised the public consciousness of identification expertise and solidified its role in the criminal justice system. Crime laboratories soon sprang up in other large cities such as Chicago and New York.¹¹ The num-

4. *People v. Jennings*, 96 N.E. 1077 (Ill. 1911).

5. See John I. Thornton, *Criminalistics: Past, Present and Future*, 11 *Lex et Scientia* 1, 23 (1975) (“In 1923, Vollmer served as Chief of Police of the City of Los Angeles for a period of one year. During that time, a crime laboratory was established at his direction.”).

6. See Federal Bureau of Investigation, U.S. Department of Justice, FBI Laboratory 3 (1981), available at <http://www.ncjrs.gov/App/publications/Abstract.aspx?id=78689>.

7. See *Anniversary Report, 40 Years of Distinguished Scientific Assistance to Law Enforcement*, FBI Law Enforcement Bull., Nov. 1972, at 4 (“During its first month of service, the FBI Laboratory examiners handled 20 cases. In its first full year of operation, the volume increased to a total of 963 examinations. By the next year that figure more than doubled.”).

8. See G. Louis Joughin & Edmund M. Morgan, *The Legacy of Sacco & Vanzetti* 15 (1948); see also James E. Starrs, *Once More Unto the Breech: The Firearms Evidence in the Sacco and Vanzetti Case Revisited*, Parts I & II, 31 *J. Forensic Sci.* 630, 1050 (1986).

9. See D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification “Expertise,”* 137 *U. Pa. L. Rev.* 731, 738 (1989).

10. See Shirley A. Graham, *Anatomy of the Lindbergh Kidnapping*, 42 *J. Forensic Sci.* 368 (1997). The kidnapper had used a wooden ladder to reach the second-story window of the child’s bedroom. Arthur Koehler, a wood technologist and identification expert for the Forest Products Laboratory of the U.S. Forest Service, traced part of the ladder’s wood from its mill source to a lumberyard near the home of the accused. Relying on plant anatomical comparisons, he also testified that a piece of the ladder came from a floorboard in the accused’s attic.

11. See Joseph L. Peterson, *The Crime Lab*, in *Thinking About Police* 184, 185 (Carl Klockars ed., 1983) (“[T]he Chicago Crime Laboratory has the distinction of being one of the oldest in the country. Soon after, however, many other jurisdictions also built police laboratories in an attempt to cope with the crimes of violence associated with the 1930s gangster era.”).

ber of laboratories gradually grew and then skyrocketed. The national campaign against drug abuse led most crime laboratories to create forensic chemistry units, and today the analysis of suspected contraband drugs constitutes more than 50% of the caseload of many laboratories.¹² By 2005, the nation's crime laboratories were handling approximately 2.7 million cases every year.¹³ According to a 2005 census, there are now 389 publicly funded crime laboratories in the United States: 210 state or regional laboratories, 84 county laboratories, 62 municipal laboratories, and 33 federal laboratories.¹⁴ Currently, these laboratories employ more than 11,900 full-time staff members.¹⁵

The establishment of crime laboratories represented a significant reform in the types of evidence used in criminal trials. Previously, prosecutors had relied primarily on eyewitness testimony and confessions. The reliability of physical evidence is often superior to that of other types of proof.¹⁶ However, the seeds of the current controversies over forensic identification expertise were sown during this period. Even though the various techniques became the stock and trade of crime laboratories, many received their judicial imprimatur without a critical evaluation of the supporting scientific research.¹⁷

This initial lack of scrutiny resulted, in part, from the deference that previous standards of admissibility accorded the community of specialists in the various fields of expert testimony. In 1923, the D.C. Circuit adopted the "general accep-

12. J. Peterson & M. Hickman, Bureau of Just. Stat. Bull. (Feb. 2005), NCJ 207205. In most cases, the forensic chemist simply identifies the unknown as a particular drug. However, in some cases the chemist attempts to individuate and establish that several drug samples originated from the same production batch at a particular illegal drug laboratory. See Fabrice Besacier et al., *Isotopic Analysis of 13C as a Tool for Comparison and Origin Assignment of Seized Heroin Samples*, 42 J. Forensic Sci. 429 (1997); C. Sten et al., *Computer Assisted Retrieval of Common-Batch Members in Leukart Amphetamine Profiling*, 38 J. Forensic Sci. 1472 (1993).

13. Matthew R. Durose, *Crime Labs Received an Estimated 2.7 Million Cases in 2005*, Bureau of Just. State. Bull. (July 2008) NCJ 222181, available at <http://pjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&lid=490> (summarizing statistics compiled by the Justice Department's Bureau of Justice Statistics).

14. NRC Forensic Science Report, *supra* note 3, at 58.

15. *Id.* at 59.

16. For example, in 1927, Justice Frankfurter, then a law professor, sharply critiqued the eyewitness identifications in the Sacco and Vanzetti case. See Felix Frankfurter, *The Case of Sacco and Vanzetti* 30 (1927) ("What is the worth of identification testimony even when uncontradicted? The identification of strangers is proverbially untrustworthy."). In 1936, the Supreme Court expressed grave reservations about the trustworthiness of confessions wrung from a suspect by abusive interrogation techniques. See *Brown v. Mississippi*, 297 U.S. 278 (1936) (due process violated by beating a confession out of a suspect).

17. "[F]ingerprints were accepted as an evidentiary tool without a great deal of scrutiny or skepticism" of their underlying assumptions. Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 Brook. L. Rev. 13, 17 (2001); see also Risinger et al., *supra* note 9, at 738 ("Our literature search for empirical evaluation of handwriting identification turned up one primitive and flawed validity study from nearly 50 years ago, one 1973 paper that raises the issue of consistency among examiners but presents only uncontrolled impressionistic and anecdotal information not qualifying as data in any rigorous sense, and a summary of one study in a 1978 government report. Beyond this, nothing.").

tance” test for determining the admissibility of scientific evidence. The case, *Frye v. United States*,¹⁸ involved a precursor of the modern polygraph. Although the general acceptance test was limited to mostly polygraph cases for several decades, it eventually became the majority pre-*Daubert* standard.¹⁹ However, under that test, scientific testimony is admissible if the underlying theory or technique is generally accepted by the specialists within the expert’s field. The *Frye* test did not require foundational proof of the empirical validity of the technique’s scientific premises.

III. Reappraisal of Forensic Identification Expertise

The advent of DNA profiling in the late 1980s, quickly followed by the Supreme Court’s 1993 *Daubert* decision (rejecting *Frye*), prompted a reassessment of identification expertise.²⁰

A. DNA Profiling and Empirical Testing

In many ways, DNA profiling revolutionized the use of expert testimony in criminal cases.²¹ Population geneticists, often affiliated with universities, used statistical techniques to define the extent to which a match of DNA markers individuated the accused as the possible source of the crime scene sample.²² Typically, the experts testified to a random-match probability, supporting their opinions by pointing to extensive empirical testing.

The fallout from the introduction of DNA analysis in criminal trials was significant in three ways. First, DNA profiling became the gold standard, regarded as the most reliable of all forensic techniques.²³ NRC issued two reports on the

18. 293 F. 1013 (D.C. Cir. 1923).

19. *Frye* was cited only five times in published opinions before World War II, mostly in polygraph cases. After World War II, it was cited 6 times before 1950, 20 times in the 1950s, and 21 times in the 1960s. Bert Black et al., *Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge*, 72 Tex. L. Rev. 715, 722 n.30 (1994).

20. See Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 Science 892 (2005).

21. See *People v. Wesley*, 533 N.Y.S.2d 643, 644 (County Ct. 1988) (calling DNA evidence the “single greatest advance in the ‘search for truth’ . . . since the advent of cross-examination”).

22. DNA Profiling is examined in detail in David H. Kaye & George Sensabaugh, Reference Guide on DNA Identification Evidence, in this manual.

23. See Michael Lynch, *God’s Signature: DNA Profiling, The New Gold Standard in Forensic Science*, 27 *Endeavour* 2, 93 (2003); Joseph L. Peterson & Anna S. Leggett, *The Evolution of Forensic Science: Progress Amid the Pitfalls*, 36 *Stetson L. Rev.* 621, 654 (2007) (“The scientific integrity and reliability of DNA testing have helped DNA replace fingerprinting and made DNA evidence the new ‘gold standard’ of forensic evidence”); see also NRC Forensic Science Report, *supra* note 3, at 40–41 (the ascendancy of DNA).

subject, emphasizing the importance of certain practices: “No laboratory should let its results with a new DNA typing method be used in court, unless it has undergone . . . proficiency testing via blind trials.”²⁴ Commentators soon pointed out the broader implications of this development:

The increased use of DNA analysis, which has undergone extensive validation, has thrown into relief the less firmly credentialed status of other forensic science identification techniques (fingerprints, fiber analysis, hair analysis, ballistics, bite marks, and tool marks). These have not undergone the type of extensive testing and verification that is the hallmark of science elsewhere.²⁵

Second, the DNA admissibility battles highlighted the absence of mandatory regulation of crime laboratories.²⁶ This situation began to change with the passage of the DNA Identification Act of 1994,²⁷ the first federal statute regulating a crime laboratory procedure. The Act authorized the creation of a national database for the DNA profiles of convicted offenders as well as a database for unidentified profiles from crime scenes: the Combined DNA Index System (CODIS). Bringing CODIS online was a major undertaking, and its successful operation required an effective quality assurance program. As one government report noted, “the integrity of the data contained in CODIS is extremely important since the DNA matches provided by CODIS are frequently a key piece of evidence linking a suspect to a crime.”²⁸ The statute also established a DNA Advisory Board (DAB) to assist in promulgating quality assurance standards²⁹ and required proficiency

24. National Research Council, *DNA Technology in Forensic Science* 55 (1992) [hereinafter NRC I], available at http://www.nap.edu/catalog.php?record_id=1866. A second report followed. See National Research Council, *The Evaluation of Forensic DNA Evidence* (1996), available at http://www.nap.edu/catalog.php?record_id=5141. The second report also recommended proficiency testing. *Id.* at 88 (Recommendation 3.2: “Laboratories should participate regularly in proficiency tests, and the results should be available for court proceedings.”).

25. Donald Kennedy & Richard A. Merrill, *Assessing Forensic Science*, 20 *Issues in Sci. & Tech.* 33, 34 (2003); see also Michael J. Saks & Jonathan J. Koehler, *What DNA “Fingerprinting” Can Teach the Law About the Rest of Forensic Science*, 13 *Cardozo L. Rev.* 361, 372 (1991) (“[F]orensic scientists, like scientists in all other fields, should subject their claims to methodologically rigorous empirical tests. The results of these tests should be published and debated.”); Sandy L. Zabell, *Fingerprint Evidence*, 13 *J.L. & Pol’y* 143, 143 (2005) (“DNA identification has not only transformed and revolutionized forensic science, it has also created a new set of standards that have raised expectations for forensic science in general.”).

26. In 1989, Eric Lander, a prominent molecular biologist who became enmeshed in the early DNA admissibility disputes, wrote: “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.” Eric S. Lander, *DNA Fingerprinting on Trial*, 339 *Nature* 501, 505 (1989).

27. 42 U.S.C. § 14131 (2004).

28. Office of Inspector General, U.S. Department of Justice, *Audit Report, The Combined DNA Index System*, ii (2001), available at <http://www.justice.gov/oig/reports/FBI/a0126/final.pdf>.

29. 42 U.S.C. § 14131(b). The legislation contained a “sunset” provision; DAB would expire after 5 years unless extended by the Director of the FBI. The board was extended for several months and then ceased to exist. The FBI had established the Technical Working Group on DNA Identifica-

testing for FBI analysts as well as those in laboratories participating in the national database or receiving federal funding.³⁰

Third, the use of DNA evidence to exonerate innocent convicts led to a reexamination of the evidence admitted to secure their original convictions.³¹ Some studies indicated that, after eyewitness testimony, forensic identification evidence was one of the most common types of testimony that jurors relied on at the earlier trials in returning erroneous verdicts.³² These studies suggested that flawed forensic analyses may have contributed to the convictions.³³

B. *Daubert and Empirical Testing*

The second major development prompting a reappraisal of forensic identification evidence was the *Daubert* decision.³⁴ Although there was some uncertainty about the effect of the decision at the time *Daubert* was decided, the Court's subsequent cases, *General Electric Co. v. Joiner*³⁵ and *Kumho Tire Co. v. Carmichael*,³⁶ signaled

tion Methods (TWGDAM) in 1988 to develop standards. TWGDAM functioned under DAB. It was renamed the Scientific Working Group on DNA Analysis Methods (SWGDM) in 1999 and replaced DAB when the latter expired.

30. 42 U.S.C. § 14132(b)(2) (2004) (external proficiency testing for CODIS participation); *id.* § 14133(a)(1)(A) (2004) (FBI examiners). DAB Standard 13 implements this requirement. The Justice for All Act, enacted in 2004, amended the statute, requiring all DNA labs to be accredited within 2 years "by a nonprofit professional association of persons actively involved in forensic science that is nationally recognized within the forensic science community" and to "undergo external audits, not less than once every 2 years, that demonstrate compliance with standards established by the Director of the Federal Bureau of Investigation." 42 U.S.C. § 14132(b)(2).

31. See Samuel R. Gross et al., *Exonerations in the United States 1989 Through 2003*, 95 J. Crim. L. & Criminology 523, 543 (2005).

32. A study of 200 DNA exonerations found that expert testimony (55%) was the second leading type of evidence (after eyewitness identifications, 79%) used in the wrongful conviction cases. Pre-DNA serology of blood and semen evidence was the most commonly used technique (79 cases). Next came hair evidence (43 cases), soil comparison (5 cases), DNA tests (3 cases), bite mark evidence (3 cases), fingerprint evidence (2 cases), dog scent (2 cases), spectrographic voice evidence (1 case), shoe prints (1 case), and fibers (1 case). Brandon L. Garrett, *Judging Innocence*, 108 Colum. L. Rev. 55, 81 (2008). These data do not necessarily mean that the forensic evidence was improperly used. For example, serological testing at the time of many of these convictions was simply not as discriminating as DNA profiling. Consequently, a person could be included using these serological tests but be excluded by DNA analysis. Yet, some evidence was clearly misused. See also Paul C. Giannelli, *Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs*, 86 N.C. L. Rev. 163, 165–70, 172–207 (2007).

33. See *Melendez-Diaz v. Massachusetts*, 129 S. Ct. 2527, 2537 (2009) (citing Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 Va. L. Rev. 1, 34–84 (2009)). See also Brandon L. Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*, ch. 4 (2011).

34. *Daubert* is discussed in detail in Margaret A. Berger, *The Admissibility of Expert Testimony*, in this manual.

35. 522 U.S. 136 (1997).

36. 526 U.S. 137 (1999).

that the *Daubert* standard may often be more demanding than the traditional *Frye* standard.³⁷ *Kumho* extended the reliability requirement to all types of expert testimony, and in 2000, the Court characterized *Daubert* as imposing an “exacting” standard for the admissibility of expert testimony.³⁸

Daubert’s impact in civil cases is well documented.³⁹ Although *Daubert*’s effect on criminal litigation has been less pronounced,⁴⁰ it nonetheless has partially changed the legal landscape. Defense attorneys invoked *Daubert* as the basis for mounting attacks on forensic identification evidence, and a number of courts view the *Daubert* trilogy as “inviting a reexamination even of ‘generally accepted’ venerable, technical fields.”⁴¹ Several courts have held that a forensic technique is not exempt from Rule 702 scrutiny simply because it previously qualified for admission under *Frye*’s general acceptance standard.⁴²

In addition to enunciating a new reliability test, *Daubert* listed several factors that trial judges may consider in assessing reliability. The first and most important *Daubert* factor is testability. Citing scientific authorities, the *Daubert* Court noted that a hallmark of science is empirical testing. The Court quoted Hempel:

37. See *United States v. Horn*, 185 F. Supp. 2d 530, 553 (D. Md. 2002) (“Under *Daubert*, . . . it was expected that it would be easier to admit evidence that was the product of new science or technology. In practice, however, it often seems as though the opposite has occurred—application of *Daubert/Kumho* Tire analysis results in the exclusion of evidence that might otherwise have been admitted under *Frye*.”).

38. *Weisgram v. Marley Co.*, 528 U.S. 440, 455 (2000).

39. See Lloyd Dixon & Brian Gill, *Changes in the Standards of Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision* 25 (2002) (“[S]ince *Daubert*, judges have examined the reliability of expert evidence more closely and have found more evidence unreliable as a result.”); Margaret A. Berger, *Upsetting the Balance Between Adverse Interests: The Impact of the Supreme Court’s Trilogy on Expert Testimony in Toxic Tort Litigation*, 64 *Law & Contemp. Probs.* 289, 290 (2001) (“The Federal Judicial Center conducted surveys in 1991 and 1998 asking federal judges and attorneys about expert testimony. In the 1991 survey, seventy-five percent of the judges reported admitting all proffered expert testimony. By 1998, only fifty-nine percent indicated that they admitted all proffered expert testimony without limitation. Furthermore, sixty-five percent of plaintiff and defendant counsel stated that judges are less likely to admit some types of expert testimony since *Daubert*.”).

40. See Jennifer L. Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 *Psychol. Pub. Pol’y & L.* 339, 364 (2002) (“[T]he *Daubert* decision did not impact on the admission rates of expert testimony at either the trial or the appellate court levels.”); D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?* 64 *Alb. L. Rev.* 99, 149 (2000) (“[T]he heightened standards of dependability imposed on expertise proffered in civil cases has continued to expand, but . . . expertise proffered by the prosecution in criminal cases has been largely insulated from any change in pre-*Daubert* standards or approach.”).

41. *United States v. Hines*, 55 F. Supp. 2d 62, 67 (D. Mass. 1999) (handwriting comparison); see also *United States v. Hidalgo*, 229 F. Supp. 2d 961, 966 (D. Ariz. 2002) (“Courts are now confronting challenges to testimony, as here, whose admissibility had long been settled”; discussing handwriting comparison).

42. See, e.g., *United States v. Williams*, 506 F.3d 151, 162 (2d Cir. 2007) (“Nor did [*Daubert*] ‘grandfather’ or protect from *Daubert* scrutiny evidence that had previously been admitted under *Frye*.”); *United States v. Starzeczyel*, 880 F. Supp. 1027, 1040 n.14 (S.D.N.Y. 1995).

“[T]he statements constituting a scientific explanation must be capable of empirical test,”⁴³ and then Popper: “[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.”⁴⁴ The other factors listed by the Court are generally complementary. For example, the second factor, peer review and publication, is a means to verify the results of the testing mentioned in the first factor; and in turn, verification can lead to general acceptance of the technique within the broader scientific community.⁴⁵ These factors serve as circumstantial evidence that other experts have examined the underlying research and found it to be sound. Similarly, another factor, an error rate, is derived from testing.

IV. National Research Council Report on Forensic Science

In 2005, the Science, State, Justice, Commerce, and Related Agencies Appropriations Act became law.⁴⁶ The accompanying Senate report commented that, “[w]hile a great deal of analysis exists of the requirements of the discipline of DNA, there exists little or no analysis of the . . . needs of the [forensic] community outside of the area of DNA.”⁴⁷ In the Act, Congress authorized the National Academy of Sciences (NAS) to conduct a comprehensive study of the current state of forensic science to develop recommendations. In fall 2006, the Academy established the Committee on Identifying the Needs of the Forensic Science Community within NRC to fulfill the task appointed by Congress. In February 2009, NRC released the report *Strengthening Forensic Science in the United States: A Path Forward*.⁴⁸

43. Carl G. Hempel, *Philosophy of Natural Science* 49 (1966).

44. Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* 37 (5th ed. 1989).

45. In their amici brief in *Daubert*, the *New England Journal of Medicine* and other medical journals observed:

“Good science” is a commonly accepted term used to describe the scientific community’s system of quality control which protects the community and those who rely upon it from unsubstantiated scientific analysis. It mandates that each proposition undergo a rigorous trilogy of publication, replication and verification before it is relied upon.

Brief for the New England Journal of Medicine, Journal of the American Medical Association, and Annals of Internal Medicine as Amici Curiae supporting Respondent at *2, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993) (No. 92-102), 1993 WL 13006387. Peer review’s “role is to promote the publication of well-conceived articles so that the most important review, the consideration of the reported results by the scientific community, may occur after publication.” *Id.* at *3.

46. Pub. L. No. 109-108, 119 Stat. 2290 (2005).

47. S. Rep. No. 109-88, at 46 (2005).

48. NRC Forensic Science Report, *supra* note 3. The Supreme Court cited the report 3 months later. *Melendez-Diaz v. Massachusetts*, 129 S. Ct. 2527 (2009).

In keeping with its congressional charge, the NRC committee did not address admissibility issues. The NRC report stated: “No judgment is made about past convictions and no view is expressed as to whether courts should reassess cases that already have been tried.”⁴⁹ When the report was released, the co-chair of the NRC committee stated:

I want to make it clear that the committee’s report does not mean to offer any judgments on any cases in the judicial system. The report does not assess past criminal convictions, nor does it speculate about pending or future cases. And the report offers no proposals for law reform. That was beyond our charge. Each case in the criminal justice system must be decided on the record before the court pursuant to the applicable law, controlling precedent, and governing rules of evidence. The question whether forensic evidence in a particular case is admissible under applicable law is not coterminous with the question whether there are studies confirming the scientific validity and reliability of a forensic science discipline.⁵⁰

Yet, in one passage, the report remarked: “Much forensic evidence—including, for example, bite marks and firearm and toolmark identifications—is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline.”⁵¹ Moreover, the report did discuss a number of forensic techniques and, where relevant, passages from the report are cited throughout this chapter.

As the NRC report explained, its primary focus is forward-looking—to outline an “agenda for progress.”⁵² The report’s recommendations are wide-ranging, covering diverse topics such as medical examiner systems,⁵³ interoperability of the automated fingerprint systems,⁵⁴ education and training in the forensic sciences,⁵⁵ codes of ethics,⁵⁶ and homeland security issues.⁵⁷ Some recommendations are

49. *Id.* at 85. The report goes on to state:

The report finds that the existing legal regime—including the rules governing the admissibility of forensic evidence, the applicable standards governing appellate review of trial court decisions, the limitations of the adversary process, and judges and lawyers who often lack the scientific expertise necessary to comprehend and evaluate forensic evidence—is inadequate to the task of curing the documented ills of the forensic science disciplines.

Id.

50. Harry T. Edwards, Co-Chair, Forensic Science Committee, Opening Statement of Press Conference (Feb. 18, 2009), transcript available at <http://www.nationalacademies.org/includes/OSEdwards.pdf>.

51. NRC Forensic Science Report, *supra* note 3, at 107–08.

52. *Id.* at xix.

53. Recommendation 10 (urging the replacement of the coroner with medical examiner system in medicolegal death investigation).

54. Recommendation 11.

55. Recommendation 2.

56. Recommendation 9.

57. Recommendation 12.

structural—that is, the creation of an independent federal entity (to be named the National Institute of Forensic Sciences) to oversee the field⁵⁸ and the removal of crime laboratories from the “administrative” control of law enforcement agencies.⁵⁹ The National Institute of Forensic Sciences would be responsible for (1) establishing and enforcing best practices for forensic science professionals and laboratories; (2) setting standards for the mandatory accreditation of crime laboratories and the mandatory certification of forensic scientists; (3) promoting scholarly, competitive, peer-reviewed research and technical development in the forensic sciences; and (4) developing a strategy to improve forensic science research. Congressional action would be needed to establish the institute. Several other recommendations are discussed below.

A. Research

The NRC report urged funding for additional research “to address issues of accuracy, reliability, and validity in the forensic science disciplines.”⁶⁰ In the report’s words, “[a]mong existing forensic methods, only nuclear DNA analysis has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between an evidentiary sample and a specific individual or source.”⁶¹ In another passage, the report discussed the need for further research into the premises underlying forensic disciplines other than DNA:

A body of research is required to establish the limits and measures of performance and to address the impact of sources of variability and potential bias. Such research is sorely needed, but it seems to be lacking in most of the forensic disciplines that rely on subjective assessments of matching characteristics. These disciplines need to develop rigorous protocols to guide these subjective interpretations and pursue equally rigorous research and evaluation programs.⁶²

58. Recommendation 1.

59. Recommendation 4.

60. *Id.* at 22 (Recommendation 3).

61. *Id.* at 100; *see also id.* at 7 & 87.

62. *Id.* at 8; *see also id.* at 15 (“Of the various facets of underresourcing, the committee is most concerned about the knowledge base. Adding more dollars and people to the enterprise might reduce case backlogs, but it will not address fundamental limitations in the capabilities of forensic science disciplines to discern valid information from crime scene evidence.”); *id.* at 22 (“[S]ome forensic science disciplines are supported by little rigorous systematic research to validate the discipline’s basic premises and techniques. There is no evident reason why such research cannot be conducted.”).

B. Observer Effects

Another recommendation focuses on research to investigate observer bias and other sources of human error in forensic examinations.⁶³ According to psychological theory of observer effects, external information provided to persons conducting analyses may taint their conclusions—a serious problem in techniques with a subjective component.⁶⁴ A growing body of modern research, noted in the report,⁶⁵ demonstrates that exposure to such information can affect forensic science experts. For example, a handwriting examiner who is informed that an exemplar belongs to the prime suspect in a case may be subconsciously influenced by this information.⁶⁶

One of the first studies to document the biasing effect was a research project involving hair analysts.⁶⁷ Some recent studies involving fingerprints have found biasing.⁶⁸ Another study concluded that external information had an effect but not toward making errors. Instead, these researchers found fewer definitive and

63. Recommendation 8:

Such programs might include studies to determine the effects of contextual bias in forensic practice (e.g., studies to determine whether and to what extent the results of forensic analyses are influenced by knowledge regarding the background of the suspect and the investigator's theory of the case). In addition, research on sources of human error should be closely linked with research conducted to quantify and characterize the amount of error.

64. See generally D. Michael Risinger et al., *The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 Cal. L. Rev. 1 (2002).

65. NRC Forensic Science Report, *supra* note 3, at 139 n.23 & 185 n.2.

66. See L.S. Miller, *Bias Among Forensic Document Examiners: A Need for Procedural Change*, 12 J. Police Sci. & Admin. 407, 410 (1984) (“The conclusions and opinions reported by the examiners supported the bias hypothesis.”). Confirmation bias is another illustration. The FBI noted the problem in its internal investigation of the Mayfield case. A review by another examiner was not conducted blind—that is, the reviewer knew that a positive identification had already been made—and thus was subject to the influence of confirmation bias. Robert B. Stacey, *A Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 54 J. Forensic Identification 707 (2004).

67. See Larry S. Miller, *Procedural Bias in Forensic Science Examinations of Human Hair*, 11 Law & Hum. Behav. 157 (1987). In the conventional method, the examiner is given hair samples from a known suspect along with a report including other facts and information relating to the guilt of the suspect. “The findings of the present study raise some concern regarding the amount of unintentional bias among human hair identification examiners. . . . A preconceived conclusion that a questioned hair sample and a known hair sample originated from the same individual may influence the examiner’s opinion when the samples are similar.” *Id.* at 161.

68. See Itiel Dror & Robert Rosenthal, *Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts*, 53 J. Forensic Sci. 900 (2008); Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 Forensic Sci. Int’l 74 (2006); Itiel Dror et al., *When Emotions Get the Better of Us: The Effect of Contextual Tap-Down Processing on Matching Fingerprints*, 19 App. Cognit. Psychol. 799 (2005).

erroneous judgments.⁶⁹ In any event, forensic examinations should, to the extent feasible, be conducted “blind.”⁷⁰

C. Accreditation and Certification

The NRC report called for the mandatory accreditation of crime labs and the certification of examiners.⁷¹ Accreditation and certification standards should be based on recognized international standards, such as those published by the International Organization for Standardization (ISO). According to the report, no person (public or private) ought to practice or testify as a forensic expert without certification.⁷² In addition, laboratories should establish “quality assurance and quality control procedures to ensure the accuracy of forensic analyses and the work of forensic practitioners.”⁷³

The American Society of Crime Lab Directors/Laboratory Accreditation Board (ASCLD/LAB) is the principal accrediting organization in the United States. Accreditation requirements generally include ensuring the integrity of evidence, adhering to valid and generally accepted procedures, employing qualified examiners, and operating quality assurance programs—that is, proficiency testing, technical reviews, audits, and corrective action procedures.⁷⁴ Currently, accreditation is mostly voluntary. Only a few states require accreditation of crime

69. Glenn Langenburg et al., *Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology When Conducting Fingerprint Comparisons*, 54 J. Forensic Sci. 571 (2009). As the researchers acknowledge, the examiners knew that they were being tested.

70. See Mike Redmayne, *Expert Evidence and Criminal Justice* 16 (2001) (“To the extent that we are aware of our vulnerability to bias, we may be able to control it. In fact, a feature of good scientific practice is the institution of processes—such as blind testing, the use of precise measurements, standardized procedures, statistical analysis—that control for bias.”).

71. Recommendation 3; see also NRC Forensic Science Report, *supra* note 3, at 23 (“In short, oversight and enforcement of operating standards, certification, accreditation, and ethics are lacking in most local and state jurisdictions.”).

72. *Id.*, Recommendation 7. The recommendation goes on to state:

Certification requirements should include, at a minimum, written examinations, supervised practice, proficiency testing, continuing education, recertification procedures, adherence to a code of ethics, and effective disciplinary procedures. All laboratories (public or private) should be accredited and all forensic science professionals should be certified, when eligible, within a time period established by NIFS.

73. *Id.*, Recommendation 8. The recommendation further comments: “Quality control procedures should be designed to: identify mistakes, fraud, and bias; confirm the continued validity and reliability of standard operating procedures and protocols; ensure that best practices are being followed; and correct procedures and protocols that are found to need improvement.”

74. See Jan S. Bashinski & Joseph L. Peterson, *Forensic Sciences, in Local Government: Police Management* 559, 578 (William Geller & Darrel Stephens eds., 4th ed. 2004).

laboratories.⁷⁵ New York mandated accreditation in 1994.⁷⁶ Texas⁷⁷ and Oklahoma⁷⁸ followed after major crime laboratory failures.

D. Proficiency Testing

Several of the report's recommendations referred to proficiency testing,⁷⁹ of which there are several types: internal or external, and blind or nonblind (declared).⁸⁰ The results of the first Laboratory Proficiency Testing Program, sponsored by the Law Enforcement Assistance Administration (LEAA), were reported in 1978.⁸¹ Voluntary proficiency testing continued after this study.⁸² The DNA Identification Act of 1994 mandated proficiency testing for examiners at the FBI as well as for

75. The same is true for certification. NRC Forensic Science Report, *supra* note 3, at 6 (“[M]ost jurisdictions do not require forensic practitioners to be certified, and most forensic science disciplines have no mandatory certification program.”).

76. N.Y. Exec. Law § 995-b (McKinney 2003) (requiring accreditation by the state Forensic Science Commission); *see also* Cal. Penal Code § 297 (West 2004) (requiring accreditation of DNA units by ASCLD/LAB or any certifying body approved by ASCLD/LAB); Minn. Stat. Ann. § 299C.156(2) (4) (West Supp. 2006) (specifying that the Forensic Science Advisory Board should encourage accreditation by ASCLD/LAB or other accrediting body).

77. Tex. Code Crim. Proc. Ann. art. 38.35 (Vernon 2004) (requiring accreditation by the Department of Public Safety). Texas also created a Forensic Science Commission. *Id.* art. 38.01 (2007).

78. Okla. Stat. Ann. tit. 74, § 150.37(D) (West 2004) (requiring accreditation by ASCLD/LAB or the American Board of Forensic Toxicology).

79. Recommendations 6 & 7.

80. Proficiency testing does not automatically correlate with a technique's “error rate.” There is a question whether error rate should be based on the results of declared and/or blind proficiency tests of simulated evidence administered to crime laboratories, or if this rate should be based on the retesting of actual case evidence drawn randomly (1) from the files of crime laboratories or (2) from evidence presented to courts in prosecuted and/or contested cases.

81. Joseph L. Peterson et al., *Crime Laboratory Proficiency Testing Research Program* (1978) [hereinafter *Laboratory Proficiency Test*]. The report concluded: “A wide range of proficiency levels among the nation's laboratories exists, with several evidence types posing serious difficulties for the laboratories. . . .” *Id.* at 3. Although the proficiency tests identified few problems in certain forensic disciplines such as glass analysis, tests of other disciplines such as hair analysis produced very high rates of “unacceptable proficiency.” According to the report, unacceptable proficiency was most often caused by (1) misinterpretation of test results due to carelessness or inexperience, (2) failure to employ adequate or appropriate methodology, (3) mislabeling or contamination of primary standards, and (4) inadequate databases or standard spectra. *Id.* at 258.

82. *See* Joseph L. Peterson & Penelope N. Markham, *Crime Laboratory Proficiency Testing Results, 1978–1991, Part I: Identification and Classification of Physical Evidence*, 40 *J. Forensic Sci.* 994 (1995); Joseph L. Peterson & Penelope N. Markham, *Crime Laboratory Proficiency Testing Results, 1978–1991, Part II: Resolving Questions of Common Origin*, 40 *J. Forensic Sci.* 1009 (1995). After collaborating with the Forensic Sciences Foundation in the initial LEAA-funded crime laboratory proficiency testing research program, Collaborative Testing Services, Inc. (CTS) began in 1978 to offer a fee-based testing program. Today, CTS offers samples in many scientific evidence testing areas to more than 500 forensic science laboratories worldwide. *See* test results at www.collaborativetesting.com/.

analysts in laboratories that participate in the national DNA database or receive federal funding.⁸³

E. Standard Terminology

The NRC report voiced concern about the use of terms such as “match,” “consistent with,” “identical,” “similar in all respects tested,” and “cannot be excluded as the source of.” These terms can have “a profound effect on how the trier of fact in a criminal or civil matter perceives and evaluates scientific evidence.”⁸⁴ Such terms need to be defined and standardized, according to the report.

F. Laboratory Reports

A related recommendation concerns laboratory reports and the need for model formats.⁸⁵ The NRC report commented:

As a general matter, laboratory reports generated as the result of a scientific analysis should be complete and thorough. They should contain, at minimum, “methods and materials,” “procedures,” “results,” “conclusions,” and, as appropriate, sources and magnitudes of uncertainty in the procedures and conclusions (e.g., levels of confidence). Some forensic science laboratory reports meet this standard of reporting, but many do not. Some reports contain only identifying and agency information, a brief description of the evidence being submitted, a brief description of the types of analysis requested, and a short statement of the results (e.g., “the greenish, brown plant material in item #1 was identified as marijuana”), and they include no mention of methods or any discussion of measurement uncertainties.⁸⁶

In addition, reports “must include clear characterizations of the limitations of the analyses, including measures of uncertainty in reported results and associated estimated probabilities where possible.”⁸⁷

83. 42 U.S.C. § 14131(c) (2005). The DNA Act authorized a study of the feasibility of blind proficiency testing; that study raised questions about the cost and practicability of this type of examination, as well as its effectiveness when compared with other methods of quality assurance such as accreditation and more stringent external case audits. Joseph L. Peterson et al., *The Feasibility of External Blind DNA Proficiency Testing. 1. Background and Findings*, 48 J. Forensic Sci. 21, 30 (2003) (“In the extreme, blind proficiency testing is possible, but fraught with problems (including costs), and it is recommended that a blind proficiency testing program be deferred for now until it is more clear how well implementation of the first two recommendations [accreditation and external case audits] are serving the same purposes as blind proficiency testing.”).

84. NRC Forensic Science Report, *supra* note 3, at 21.

85. *Id.* at 22, Recommendation 2.

86. *Id.* at 21.

87. *Id.* at 21–22.

V. Specific Techniques

The broad field of forensic science includes disparate disciplines such as forensic pathology, forensic anthropology, arson investigation, and gunshot residue testing.⁸⁸ The NRC report explained:

Some of the forensic science disciplines are laboratory based (e.g., nuclear and mitochondrial DNA analysis, toxicology and drug analysis); others are based on expert interpretation of observed patterns (e.g., fingerprints, writing samples, toolmarks, bite marks, and specimens such as hair). . . . There are also sharp distinctions between forensic practitioners who have been trained in chemistry, biochemistry, biology, and medicine (and who bring these disciplines to bear in their work) and technicians who lend support to forensic science enterprises.⁸⁹

The report devoted special attention to forensic disciplines in which the expert's final decision is subjective in nature: "In terms of scientific basis, the analytically based disciplines generally hold a notable edge over disciplines based on expert interpretation."⁹⁰ Moreover, many of the subjective techniques attempt to render the most specific conclusions—that is, opinions concerning "individualization."⁹¹ Following the report's example, the remainder of this chapter focuses on "pattern recognition" disciplines, each of which contains a subjective component. These disciplines exemplify most of the issues that a trial judge may encounter in ruling on the admissibility of forensic testimony. Each part describes the technique, the available empirical research, and contemporary case law.

A. Terminology

Although courts often use the terms "validity" and "reliability" interchangeably, the terms have distinct meanings in scientific disciplines. "Validity" refers to the ability of a test to measure what it is supposed to measure—its accuracy. "Reliability" refers to whether the same results are obtained in each instance in which the test is performed—its consistency. Validity includes reliability, but the converse is not necessarily true. Thus, a reliable, invalid technique will consistently

88. Other examples include drug analysis, blood spatter examinations, fiber comparisons, toxicology, entomology, voice spectrometry, and explosives and bomb residue analysis. As the Supreme Court noted in *Melendez-Diaz v. Massachusetts*, 129 S. Ct. 2527, 2537–38 (2009), errors can be made when instrumental techniques, such as gas chromatography/mass spectrometry analysis, are used.

89. NRC Forensic Science Report, *supra* note 3, at 7.

90. *Id.*

91. "Often in criminal prosecutions and civil litigation, forensic evidence is offered to support conclusions about 'individualization' (sometimes referred to as 'matching' a specimen to a particular individual or other source) or about classification of the source of the specimen into one of several categories. With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source." *Id.*

yield inaccurate results. The Supreme Court acknowledged this distinction in *Daubert*, but the Court indicated that it was using the term “reliability” in a different sense. The Court wrote that its concern was “evidentiary reliability—that is, trustworthiness. . . . In a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*.”⁹²

In forensic science, class and individual characteristics are distinguished. Class characteristics are shared by a group of persons or objects (e.g., ABO blood types).⁹³ Individual characteristics are unique to an object or person. The term “match” is ambiguous because it is sometimes used to indicate the “matching” of individual characteristics, but on other occasions it is used to refer to “matching” class characteristics (e.g., blood type A at a crime scene “matches” suspect’s type A blood). Expert opinions involving “individual” and “class” characteristics raise different issues. In the former, the question is whether an individuation determination rests on a firm scientific foundation.⁹⁴ For the latter, the question is determining the size of the class.⁹⁵

VI. Fingerprint Evidence

Sir William Herschel, an Englishman serving in the Indian civil service, and Henry Faulds, a Scottish physician serving as a missionary in Japan, were among the first to suggest the use of fingerprints as a means of personal identification. Since 1858, Herschel had been collecting the handprints of natives for that purpose. In 1880, Faulds published an article entitled “On the Skin—Furrows

92. 509 U.S. at 590 n.9 (“We note that scientists typically distinguish between ‘validity’ (does the principle support what it purports to show?) and ‘reliability’ (does application of the principle produce consistent results?). . . .”).

93. See Bashinski & Peterson, *supra* note 74, at 566 (“The forensic scientist first investigates whether items possess similar ‘class’ characteristics—that is, whether they possess features shared by all objects or materials in a single class or category. (For firearms evidence, bullets of the same caliber, bearing rifling marks of the same number, width, and direction of twist, share class characteristics. They are consistent with being fired from the same *type* of weapon.) The forensic scientist then attempts to determine an item’s ‘individuality’—the features that make one thing different from all others similar to it, including those with similar class characteristics.”).

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

95. See Margaret A. Berger, *Procedural Paradigms for Applying the Daubert Test*, 78 Minn. L. Rev. 1345, 1356–57 (1994) (“We allow eyewitnesses to testify that the person fleeing the scene wore a yellow jacket and permit proof that a defendant owned a yellow jacket without establishing the background rate of yellow jackets in the community. Jurors understand, however, that others than the accused own yellow jackets. When experts testify about samples matching in every respect, the jurors may be oblivious to the probability concerns if no background rate is offered, or may be unduly prejudiced or confused if the probability of a match is confused with the probability of guilt, or if a background rate is offered that does not have an adequate scientific foundation.”).

of the Hand” in *Nature*.⁹⁶ Sir Francis Galton authored the first textbook on the subject.⁹⁷ Individual ridge characteristics came to be known as “Galton details.”⁹⁸ Subsequently, Edward Henry, the Inspector General of Police in Bengal, realized the potential of fingerprinting for law enforcement and helped establish the Fingerprint Branch at Scotland Yard when he was recalled to England in 1901.⁹⁹

English and American courts have accepted fingerprint identification testimony for just over a century. “The first English appellate endorsement of fingerprint identification testimony was the 1906 opinion in *Rex v. Castleton*. . . . In 1906 and 1908, Sergeant Joseph Fautot, a New York City detective who had in 1904 been posted to Scotland Yard to learn about fingerprinting, used his new training to break open two celebrated cases: in each instance fingerprint identification led the suspect to confess. . . .”¹⁰⁰ A 1911 Illinois Supreme Court decision, *People v. Jennings*,¹⁰¹ is the first published American appellate opinion sustaining the admission of fingerprint testimony.

Over the years, fingerprint analysis became the gold standard of forensic identification expertise. In fact, proponents of new, emerging techniques in forensics would sometimes attempt to invoke onto the new techniques the prestige of fingerprint analysis. Thus, advocates of sound spectrography referred to it as “voiceprint” analysis.¹⁰² Likewise, some early proponents of DNA typing alluded to it as “DNA fingerprinting.”¹⁰³ However, as previously noted, DNA analysis has replaced fingerprint analysis as the gold standard.

A. The Technique

Even a cursory study of fingerprints establishes that there is “intense variability . . . in even small areas of prints.”¹⁰⁴ Given that variability, it is generally assumed that an identification is possible if the comparison involves two sets of clear images of all 10 fingerprints. These are known as “record” prints and are typically rolled onto a fingerprint card or digitized and scanned into an electronic file. Two complete fingerprint sets are available for comparison in some settings such as

96. Henry Faulds, *On the Skin—Furrows of the Hand*, 22 *Nature* 605 (1881). See generally Simon Cole, *Suspect Identities: A History of Fingerprint and Criminal Identification* (2001).

97. Francis Galton, *Fingerprints* (1892).

98. See Andre A. Moenssens, *Scientific Evidence in Civil and Criminal Cases* § 10.02, at 621 (5th ed. 2007).

99. *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 554 (E.D. Pa. 2002).

100. *Id.* at 572.

101. 96 N.E. 1077 (Ill. 1911).

102. Kenneth Thomas, *Voiceprint—Myth or Miracle*, in *Scientific and Expert Evidence* 1015 (2d ed. 1981).

103. Colin Norman, *Maine Case Deals Blow to DNA Fingerprinting*, 246 *Science* 1556 (Dec. 22, 1989).

104. David A. Stoney, *Scientific Status*, in 4 David L. Faigman et al., *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 32:45, at 361 (2007–2008 ed.).

immigration matters. However, in the law enforcement setting, the task is more challenging because only a partial impression (latent print) of a single finger may be left by a criminal.

Fingerprint evidence is based on three assumptions: (1) the uniqueness of each person's friction ridges, (2) the permanence of those ridges throughout a person's life, and (3) the transferability of an impression of that uniqueness to another surface. The last point raises the most significant issue of reliability because a crime scene (latent) impression is often only a fifth of the size of the record print. Furthermore, variations in pressure and skin elasticity almost inevitably distort the impression.¹⁰⁵ Consequently, fingerprint impressions from the same person typically differ in some respects each time the impression is left on an object.¹⁰⁶

Although fingerprint analysis is based on physical characteristics, the final step in the analysis—the formation of an opinion regarding individuation—is subjective.¹⁰⁷ Examiners lack population frequency data to quantify how rare or common a particular type of fingerprint characteristic is.¹⁰⁸ Rather, in making that judgment, the examiner relies on personal experience and discussions with colleagues. Although examiners in some countries must find a certain minimum number of points of similarities between the latent and the known before declaring a match,¹⁰⁹ neither the FBI nor New Scotland Yard requires any set number.¹¹⁰ A single inexplicable difference between the two impressions precludes finding a match. Because there are frequently “dissimilarities” between the crime scene and record prints, the examiner must decide whether there is a *true* dis-

105. See *United States v. Mitchell*, 365 F.3d 215, 220–21 (3d Cir. 2004) (“Criminals generally do not leave behind full fingerprints on clean, flat surfaces. Rather, they leave fragments that are often distorted or marred by artifacts. . . . Testimony at the *Daubert* hearing suggested that the typical latent print is a fraction—perhaps 1/5th—of the size of a full fingerprint.”). “In the jargon, artifacts are generally small amounts of dirt or grease that masquerade as parts of the ridge impressions seen in a fingerprint, while distortions are produced by smudging or too much pressure in making the print, which tends to flatten the ridges on the finger and obscure their detail.” *Id.* at 221 n.1.

106. NRC Forensic Science Report, *supra* note 3, at 144 (“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.”).

107. See *Commonwealth v. Patterson*, 840 N.E.2d 12, 15, 16–17 (Mass. 2005) (“These latent print impressions are almost always partial and may be distorted due to less than full, static contact with the object and to debris covering or altering the latent impression”; “In the evaluation stage, . . . the examiner relies on his subjective judgment to determine whether the quality and quantity of those similarities are sufficient to make an identification, an exclusion, or neither”); Zbell, *supra* note 25, at 158 (“In contrast to the scientifically-based statistical calculations performed by a forensic scientist in analyzing DNA profile frequencies, each fingerprint examiner renders an opinion as to the similarity of friction ridge detail based on his subjective judgment.”).

108. NRC Forensic Science Report, *supra* note 3, at 139–40 & 144.

109. Stoney, *supra* note 104, § 32:34, at 354–55.

110. *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 566–71 (E.D. Pa. 2002).

similarity, or whether the apparent dissimilarity can be discounted as an artifact or resulting from distortion.¹¹¹

Three levels of details may be scrutinized: Level 1 details are general flow ridge patterns such as whorls, loops, and arches.¹¹² Level 2 details are fine ridges or minutiae such as bifurcations, dots, islands, and ridge endings.¹¹³ These minutiae are essentially ridge discontinuities.¹¹⁴ Level 3 details are “microscopic ridge attributes such as the width of a ridge, the shape of its edge, or the presence of a sweat pore near a particular ridge.”¹¹⁵ Within the fingerprint community there is disagreement about the usefulness and reliability of Level 3 details.¹¹⁶

FBI examiners generally follow a procedure known as analysis, comparison, evaluation, and verification (ACE-V). In the *analysis* stage, the examiner studies the latent print to determine whether the quantity and quality of details in the print are sufficient to permit further evaluation.¹¹⁷ The latent print may be so fragmentary or smudged that analysis is impossible. In the *evaluation* stage, the examiner considers at least the Level 2 details, including “the type of minutiae (forks or ridge endings), their direction (loss or production of a ridge) and their relative position (how many intervening ridges there are between minutiae and how far along the ridges it is from one minutiae to the next).”¹¹⁸ Again, if the examiner finds a single, inexplicable difference between the two prints, the examiner concludes that there is no match.¹¹⁹ Alternatively, if the examiner concludes that there is a match, the examiner seeks *verification* by a second examiner. “[T]he friction ridge community actively discourages its members from testifying in terms of the probability of a match; when a latent print examiner testifies that two impressions

111. *Patterson*, 840 N.E.2d at 17 (“There is a rule of examination, the ‘one-discrepancy’ rule, that provides that a nonidentification finding should be made if a single discrepancy exists. However, the examiner has the discretion to ignore a possible discrepancy if he concludes, based on his experience and the application of various factors, that the discrepancy might have been caused by distortions of the fingerprint at the time it was made or at the time it was collected.”).

112. *See id.* at 16 (“Level one detail involves the general ridge flow of a fingerprint, that is, the pattern of loops, arches, and whorls visible to the naked eye. The examiner compares this information to the exemplar print in an attempt to exclude a print that has very clear dissimilarities.”).

113. *See id.* (“Level two details include ridge characteristics (or Galton Points) like islands, dots, and forks, formed as the ridges begin, end, join or bifurcate.”). *See generally* FBI, *The Science of Fingerprints* (1977).

114. Stoney, *supra* note 104, § 32:31, at 350.

115. *See Patterson*, 840 N.E.2d at 16.

116. *See* Office of the Inspector General, U.S. Dep’t of Justice, *A Review of the FBI’s Handling of the Brandon Mayfield Case*, Unclassified Executive Summary 8 (Jan. 2006) *available at* www.justice.gov/oig/special/s0601/PDF_list.htm. (“Because Level 3 details are so small, the appearance of such details in fingerprints is highly variable, even between different fingerprints made by the same finger. As a result, the reliability of Level 3 details is the subject of some controversy within the latent fingerprint community.”).

117. NRC Forensic Science Report, *supra* note 3, at 137–38.

118. Stoney, *supra* note 104, § 32:31, at 350–51.

119. NRC Forensic Science Report, *supra* note 3, at 140.

‘match,’ they are communicating the notion that the prints could not possibly have come from two different individuals.”¹²⁰ The typical fingerprint analyst will give one of only three opinions: (1) the prints are unsuitable for analysis, (2) the suspect is definitely excluded, or (3) the latent print is definitely that of the suspect.

B. *The Empirical Record*

At several points, the 2009 NRC report noted that there is room for human error in fingerprint analysis. For example, the report stated that because “the ACE-V method does not specify particular measurements or a standard test protocol, . . . examiners must make subjective assessments throughout.”¹²¹ The report further commented that the ACE-V method is too “broadly stated” to “qualify as a validated method for this type of analysis.”¹²² The report added that “[t]he latent print community in the United States has eschewed numerical scores and corresponding thresholds” and consequently relies “on primarily subjective criteria” in making the ultimate attribution decision.¹²³ In making the decision, the examiner must draw on his or her personal experience to evaluate such factors as “inevitable variations” in pressure, but to date these factors have not been “characterized, quantified, or compared.”¹²⁴ At the conclusion of the section devoted to fingerprint analysis, the report outlined an agenda for the research it considered necessary “[t]o properly underpin the process of friction ridge identification.”¹²⁵ The report noted that some of these research projects have already begun.¹²⁶

Fingerprint analysis raises a number of scientific issues. For example, do the salient features of fingerprints remain constant throughout a person’s life?¹²⁷ Few of the underlying scientific premises have been subjected to rigorous empirical investigation,¹²⁸ although some experiments have been conducted, and proficiency test results are available.

Two experimental studies were discussed at the 2000 trial in *United States v. Mitchell*¹²⁹:

One of the studies conducted by the government for the *Daubert* hearing [in *Mitchell*] employed the two actual latent and the known prints that were at issue in the case. These prints were submitted to 53 state law enforcement agency

120. *Id.* at 140–41.

121. *Id.* at 139.

122. *Id.* at 142.

123. *Id.* at 141.

124. *Id.* at 144.

125. *Id.*

126. *Id.*

127. Stoney, *supra* note 104, § 32:21, at 342.

128. See Zabell, *supra* note 25, at 164 (“Although there is a substantial literature on the uniqueness of fingerprints, it is surprising how little true scientific support for the proposition exists.”).

129. 365 F.3d 215 (3d Cir. 2004).

crime laboratories around the country for their evaluation. Though, of the 35 that responded, most concluded that the latent and known prints matched, eight said that no match could be made to one of the prints and six said that no match could be made to the other print.¹³⁰

Although there were no false positives, a significant percentage of the participating laboratories reported at best inconclusive findings.

Lockheed-Martin conducted the second test, the FBI-sponsored 50K study. This was an empirical study of 50,000 fingerprint images taken from the FBI's Automated Fingerprint System, a computer database. The study

was an effort to obtain an estimate of the probability that one person's fingerprints would be mistaken for those of another person, at least to a computer system designed to match fingerprints. The FBI asked Lockheed-Martin, the manufacturer of its . . . automated fingerprint identification system, . . . to help it run a comparison of the images of 50,000 single fingerprints against the same 50,000 images, and produce a similarity score for each comparison. The point of this exercise was to show that the similarity score for an image matched against itself was far higher than the scores obtained when it was compared to the others.¹³¹

The comparisons between the two identical images yielded "extremely high scores."¹³² Nonetheless, some commentators disputed whether the Lockheed-Martin study demonstrated the validity of fingerprint analysis.¹³³ The study compared a computerized image of a fingerprint impression against other computerized images in the database. The study did not address the problem examiners encounter in the real world; it did not attempt to match a partial fingerprint impression against images in the database. As noted earlier, crime scene prints are typically distorted from pressure and sometimes only one-fifth the size of record prints.¹³⁴ Even the same finger will not leave the exact impression each time: "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."¹³⁵ Thus, one scholar asserted that the "study addresses the irrelevant question of whether one image of a fingerprint is immensely more similar to itself than to other images—including those of the same finger."¹³⁶ Citing

130. Stoney, *supra* note 104, § 32:3, at 287.

131. *Id.* § 32:3, at 288.

132. *Id.* (quoting James L. Wayman, Director, U.S. National Biometric Test Center at the College of Engineering, San Jose State University).

133. *E.g.*, David H. Kaye, *Questioning a Courtroom Proof of the Uniqueness of Fingerprints*, 71 *Int'l Statistical Rev.* 521 (2003); S. Pankanti et al., *On the Individuality of Fingerprints*, 24 *IEEE Trans. Pattern Analysis Mach. Intelligence* 1010 (2002).

134. *See supra* note 105 & accompanying text.

135. NRC Forensic Science Report, *supra* note 3, at 144.

136. Kaye, *supra* note 133, at 527–28. In another passage, he wrote: "[T]he study merely demonstrates the trivial fact that the same two-dimensional representation of the surface of a finger is far

this assertion, the 2009 NRC report stated that the Lockheed-Martin study “has several major design and analysis flaws.”¹³⁷

1. Proficiency testing

In *United States v. Llera Plaza*,¹³⁸ the district court described internal and external proficiency tests of FBI fingerprint analysts and their supervisors. Between 1995 and 2001, the supervisors participated in 16 external tests created by CTS.¹³⁹ One false-positive result was reported among the 16 tests.¹⁴⁰ During the same period, there was a total of 431 internal tests of FBI fingerprint personnel. These personnel committed no false-positive errors, but there were three false eliminations.¹⁴¹ Hence, the overall error rate was approximately 0.8%.¹⁴²

Although these proficiency tests yielded impressive accuracy rates, the quality of the tests became an issue. First, the examinees participating in the tests knew that they were being tested and, for that reason, may have been more meticulous than in regular practice. Second, the rigor of proficiency testing was questioned. The *Llera Plaza* court concluded that the FBI’s internal proficiency tests were “less demanding than they should be.”¹⁴³ In the judge’s words, “the FBI examiners got very high proficiency grades, but the tests they took did not.”¹⁴⁴

more similar to itself than to such representation of the source of finger from any other person in the data set.” *Id.* at 527.

137. NRC Forensic Science Report, *supra* note 3, at 144 n.35.

138. 188 F. Supp. 2d 549 (E.D. Pa. 2002).

139. *Id.* at 556.

140. However, a later inquiry led Stephen Meagher, Unit Chief of Latent Print Unit 3 of the Forensic Analysis Section of the FBI Laboratory “to conclude that the error was not one of faulty evaluation but of faulty recording of the evaluation—i.e., a clerical error rather than a technical error.” *Id.*

141. *Id.*

142. Sharon Begley, *Fingerprint Matches Come Under More Fire as Potentially Fallible*, Wall St. J., Oct. 7, 2005, at B1.

143. *Llera Plaza*, 188 F. Supp. 2d at 565. A fingerprint examiner from New Scotland Yard with 25 years’ experience testified that the FBI tests were deficient:

Mr. Bayle had reviewed copies of the internal FBI proficiency tests. . . . He found the latent prints utilized in those tests to be, on the whole, markedly unrepresentative of the latent prints that would be lifted at a crime scene. In general, Mr. Bayle found the test latent prints to be far clearer than the prints an examiner would routinely deal with. The prints were too clear—they were, according to Mr. Bayle, lacking in the “background noise” and “distortion” one would expect in latent prints lifted at a crime scene. Further, Mr. Bayle testified, the test materials were deficient in that there were too few latent prints that were not identifiable; according to Mr. Bayle, at a typical crime scene only about ten percent of the lifted latent prints will turn out to be matched. In Mr. Bayle’s view the paucity of non-identifiable latent prints “makes the test too easy. It’s not testing their ability. . . . [I]f I gave my experts these tests, they’d fall about laughing.”

Id. at 557–58.

144. *Id.* at 565; *see also* *United States v. Crisp*, 324 F.3d 261, 274 (4th Cir. 2003) (Michael, J., dissenting) (“Proficiency testing is typically based on a study of prints that are far superior to those usually retrieved from a crime scene.”).

In an earlier proficiency study (1995), the examiners did not do as well,¹⁴⁵ although many of the subjects were not certified FBI examiners. Of the 156 examiners who participated, only 44% reached the correct conclusion on all the identification tasks. Eighty-eight examiners or 56% provided divergent (wrong, incorrect, erroneous) answers. Six examiners failed to identify any of the latent prints. Forty eight of the 156 examiners made erroneous identifications—representing 22% of the total identifications made by the examiners.

A 2006 study resurrected some of the questions raised by the 1995 test. In that study, examiners were presented with sets of prints that they had previously reviewed.¹⁴⁶ The researchers found that “experienced examiners do not necessarily agree with even their own past conclusions when the examination is presented in a different context some time later.”¹⁴⁷

These studies call into question the soundness of testimonial claims that fingerprint analysis is infallible¹⁴⁸ or has a zero error rate.¹⁴⁹ In 2008, Haber and Haber reviewed the literature describing the ACE-V technique and the supporting research.¹⁵⁰ Although many practitioners professed using the technique, Haber and Haber found that the practitioners’ “descriptions [of their technique] differ, no single protocol has been officially accepted by the profession and the standards upon which the method’s conclusion rest[s] have not been specified quantitatively.”¹⁵¹ After considering the Haber study, NRC concluded that the ACE-V “framework is not specific enough to qualify as a validated method for this type of analysis.”¹⁵²

2. *The Mayfield case*

Like the empirical data, several reports of fingerprint misidentifications raised questions about the reliability of fingerprint analysis. The FBI misidentified Brandon Mayfield as the source of the crime scene prints in the terrorist train bombing in Madrid, Spain, on March 11, 2004.¹⁵³ The mistake was attributed in part to several types of cognitive bias. According to an FBI review, the “power” of the automated

145. See David L. Grieve, *Possession of Truth*, 46 J. Forensic Identification 521, 524–25 (1996); James Starrs, *Forensic Science on the Ropes: An Upper Cut to Fingerprinting*, 20 Sci. Sleuthing Rev. 1 (1996).

146. Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 Forensic Sci. Int’l 74, 76 (2006) (Four of five examiners changed their opinions; three directly contradicted their prior identifications, and the fourth concluded that data were insufficient to reach a definite conclusion); see also I. E. Dror & D. Charlton, *Why Experts Make Errors*, 56 J. Forensic Identification 600 (2006).

147. NRC Forensic Science Report, *supra* note 3, at 139.

148. *Id.* at 104.

149. *Id.* at 143–44.

150. Lyn Haber & Ralph Norman Haber, *Scientific Validation of Fingerprint Evidence Under Daubert*, 7 Law, Probability & Risk 87 (2008).

151. NRC Forensic Science Report, *supra* note 3, at 143.

152. *Id.* at 142.

153. *Id.* at 46 & 105.

fingerprint correlation “was thought to have influenced the examiner’s initial judgment and subsequent examination.”¹⁵⁴ Thus, he was subject to confirmation bias. Moreover, a second review by another examiner was not conducted blind—that is, the reviewer knew that a positive identification had already been made and was thus subject to expectation (context) bias. Indeed, a third expert from outside the FBI, one appointed by the court, also erroneously confirmed the identification.¹⁵⁵ In addition to the Bureau’s review, the Inspector General of the Department of Justice investigated the case.¹⁵⁶ The Mayfield case is not an isolated incident.¹⁵⁷

The Mayfield case led to a more extensive FBI review of the scientific basis of fingerprints.¹⁵⁸ In January 2006, the FBI created a three-person review committee to evaluate the fundamental basis of fingerprint analysis. The committee identified two possible approaches. One approach would be to “develop a quantifiable minimum threshold based on objective criteria”—if possible.¹⁵⁹ “Any minimum threshold must consider both the clarity (quality) and the quantity of features and include all levels of detail, not simply points or minutiae.”¹⁶⁰ Apparently, some FBI examiners use an unofficial seven-point cutoff, but this standard has never been tested.¹⁶¹ As the FBI Review cautioned: “It is compelling to focus on a quantifiable threshold; however, quality/clarity, that is, distortion and degradation of prints, is the fundamental issue that needs to be addressed.”¹⁶²

154. Stacey, *supra* note 66, at 713.

155. In addition, the culture at the laboratory was poorly suited to detect mistakes: “To disagree was not an expected response.” *Id.*

156. See Office of the Inspector General, U.S. Dep’t of Justice, *A Review of the FBI’s Handling of the Brandon Mayfield Case*, Unclassified Executive Summary 9 (Jan. 2006). The I.G. made several recommendations that went beyond the FBI’s internal report:

These include recommendations that the Laboratory [1] develop criteria for the use of Level 3 details to support identifications, [2] clarify the “one discrepancy rule” to assure that it is applied in a manner consistent with the level of certainty claimed for latent fingerprint identifications, [3] require documentation of features observed in the latent fingerprint before the comparison phase to help prevent circular reasoning, [4] adopt alternate procedures for blind verifications, [5] review prior cases in which the identification of a criminal suspect was made on the basis of only one latent fingerprint searched through IAFIS, and [6] require more meaningful and independent documentation of the causes of errors as part of the Laboratory’s corrective action procedures.

157. In 2005, Professor Cole released an article identifying 23 cases of documented fingerprint misidentifications. See Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. Crim. L. & Criminology 985 (2005). The misidentification cases include some that involved (1) verification by one or more other examiners, (2) examiners certified by the International Association of Identification, (3) procedures using a 16-point standard, and (4) defense experts who corroborated misidentifications made by prosecution experts.

158. See Bruce Budowle et al., *Review of the Scientific Basis for Friction Ridge Comparisons as a Means of Identification: Committee Findings and Recommendations*, 8 Forensic Sci. Comm. (Jan. 2006) [hereinafter *FBI Review*].

159. *Id.* at 5.

160. *Id.*

161. There is also a 12-point cutoff, under which a supervisor’s approval is required.

162. *Id.*

The second approach would treat the examiner as a “black box.” This methodology would be necessary if minimum criteria for rendering an identification cannot be devised—in other words, there is simply too much subjectivity in the process to formulate meaningful, quantitative guidelines. Under this approach, it becomes critical to determine just how good a “black box” each examiner is: “The examiner(s) can be tested with various inputs of a range of defined categories of prints. This approach would demonstrate whether or not it is possible to obtain a degree of accuracy (that is, assess the performance of the black-box examiner for rendering an identification).”¹⁶³ The review committee noted that this approach would provide the greatest assurance of reliability if it incorporated blind technical review. According to the review committee’s report, “[t]o be truly blind, the second examiner should have no knowledge of the interpretation by the first examiner (to include not seeing notes or reports).”¹⁶⁴

Although the FBI Review concluded that reliable identifications could be made, it conceded that “there are scientific areas where improvements in the practice can be made particularly regarding validation, more objective criteria for certain aspects of the ACE-V process, and data collection.”¹⁶⁵ Efforts to improve fingerprint analysis appear to be under way. In 2008, a symposium on validity testing of fingerprint examinations was published.¹⁶⁶ In late 2008, the National Institute of Standards and Technology formed the Expert Group on Human Factors in Latent Print Analysis tasked to identify the major sources of human error in fingerprint examination and to develop strategies to minimize such errors.

C. Case Law Development

As noted earlier, the seminal American decision is the Illinois Supreme Court’s 1911 opinion in *Jennings*.¹⁶⁷ Fingerprint testimony was routinely admitted in later

163. *Id.* at 4.

164. *Id.*

165. *Id.* at 10.

166. The lead article is Lyn Haber & Ralph Norman Haber, *supra* note 150. Other contributors are Christopher Champod, *Fingerprint Examination: Towards More Transparency*, 7 *Law, Probability & Risk* 111 (2008); Simon A. Cole, *Comment on “Scientific Validation of Fingerprint Evidence Under Daubert,”* 7 *Law, Probability & Risk* 119 (2008); Jennifer Mnookin, *The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate*, 7 *Law, Probability & Risk* 127 (2008).

167. *People v. Jennings*, 96 N.E. 1077 (Ill. 1911); see Donald Campbell, *Fingerprints: A Review*, [1985] *Crim. L. Rev.* 195, 196 (“Galton gave evidence to the effect that the chance of agreement would be in the region of 1 in 64,000,000,000.”). As Professor Mnookin has noted, however, “fingerprints were accepted as an evidentiary tool without a great deal of scrutiny or skepticism.” Mnookin, *supra* note 17, at 17. She elaborated:

Even if no two people had identical *sets* of fingerprints, this did not establish that no two people could have a *single* identical print, much less an identical *part* of a print. These are necessarily matters of prob-

years. Some courts stated that fingerprint evidence was the strongest proof of a person's identity.¹⁶⁸

With the exception of one federal district court decision that was later withdrawn,¹⁶⁹ the post-*Daubert* federal cases have continued to accept fingerprint testimony about individuation at least as sufficiently reliable nonscientific expertise.¹⁷⁰

Two subsequent state court decisions also deserve mention. In one, a Maryland trial judge excluded fingerprint evidence under the *Frye* test, which still controls in that state.¹⁷¹ In the other case, *Commonwealth v. Patterson*,¹⁷² the Supreme Judicial

ability, but neither the court in *Jennings* nor subsequent judges ever required that fingerprint identification be placed on a secure statistical foundation.

Id. at 19.

168. *People v. Adamson*, 165 P.2d 3, 12 (Cal. 1946), *aff'd*, 332 U.S. 46 (1947).

169. *United States v. Llera Plaza*, 179 F. Supp. 2d 492 (E.D. Pa.), *vacated, mot. granted on recons.*, 188 F. Supp. 2d 549 (E.D. Pa. 2002). The ruling was limited to excluding expert testimony that two sets of prints “matched”—that is, a positive identification to the exclusion of all other persons:

Accordingly, this court will permit the government to present testimony by fingerprint examiners who, suitably qualified as “expert” examiners by virtue of training and experience, may (1) describe how the rolled and latent fingerprints at issue in this case were obtained, (2) identify and place before the jury the fingerprints and such magnifications thereof as may be required to show minute details, and (3) point out observed similarities (and differences) between any latent print and any rolled print the government contends are attributable to the same person. What such expert witnesses will not be permitted to do is to present “evaluation” testimony as to their “opinion” (Rule 702) that a particular latent print is in fact the print of a particular person.

Id. at 516. On rehearing, however, the court reversed itself. A spate of legal articles followed. *See, e.g.,* Simon A. Cole, *Grandfathering Evidence: Fingerprint Admissibility Rulings from Jennings to Llera Plaza and Back Again*, 41 Am. Crim. L. Rev. 1189 (2004); Robert Epstein, *Fingerprints Meet Daubert: The Myth of Fingerprint “Science” Is Revealed*, 75 S. Cal. L. Rev. 605 (2002); Kristin Romandetti, *Recognizing and Responding to a Problem with the Admissibility of Fingerprint Evidence Under Daubert*, 45 *Jurimetrics J.* 41 (2004).

170. *See, e.g.,* *United States v. Baines*, 573 F.3d 979, 990 (10th Cir. 2009) (“[U]nquestionably the technique has been subject to testing, albeit less rigorous than a scientific ideal, in the world of criminal investigation, court proceedings, and other practical applications, such as identification of victims of disasters. Thus, while we must agree with defendant that this record does not show that the technique has been subject to testing that would meet all of the standards of science, it would be unrealistic in the extreme for us to ignore the countervailing evidence. Fingerprint identification has been used extensively by law enforcement agencies all over the world for almost a century.”); *United States v. Abreu*, 406 F.3d 1304, 1307 (11th Cir. 2005) (“We agree with the decisions of our sister circuits and hold that the fingerprint evidence admitted in this case satisfied *Daubert*.”); *United States v. Janis*, 387 F.3d 682, 690 (8th Cir. 2004) (finding fingerprint evidence to be reliable); *United States v. Mitchell*, 365 F.3d 215, 234–52 (3d Cir. 2004); *United States v. Crisp*, 324 F.3d 261, 268–71 (4th Cir. 2003); *United States v. Collins*, 340 F.3d 672, 682 (8th Cir. 2002) (“Fingerprint evidence and analysis is generally accepted.”); *United States v. Hernandez*, 299 F.3d 984, 991 (8th Cir. 2002); *United States v. Sullivan*, 246 F. Supp. 2d 700, 704 (E.D. Ky. 2003); *United States v. Martinez-Cintron*, 136 F. Supp. 2d 17, 20 (D.P.R. 2001).

171. *State v. Rose*, No. K06-0545, 2007 WL 5877145 (Cir. Ct. Baltimore, Md., Oct. 19, 2007). *See* NRC Forensic Science Report, *supra* note 3, at 43 & 105. However, in a parallel federal case, the evidence was admitted. *United States v. Rose*, 672 F. Supp. 2d 723 (D. Md. 2009).

172. 840 N.E.2d 12 (Mass. 2005).

Court of Massachusetts considered the reliability of applying the ACE-V methodology to simultaneous impressions. Simultaneous impressions “are two or more friction ridge impressions from the fingers and/or palm on one hand that are determined to have been deposited at the same time.”¹⁷³ The key is deciding whether the impressions were left at the same time and therefore came from the same person, rather than having been left by two different people at different times.¹⁷⁴ Although the court found that the ACE-V method is generally accepted by the relevant scientific community, the record did not demonstrate similar acceptance of that methodology as applied to simultaneous impressions. The court consequently remanded the case to the trial court.¹⁷⁵

VII. Handwriting Evidence

The Lindbergh kidnapping trial showcased testimony by questioned document examiners. Later, in the litigation over Howard Hughes’ alleged will, both sides relied on handwriting comparison experts.¹⁷⁶ Thanks in part to such cases, questioned document examination expertise has enjoyed widespread use and judicial acceptance.

A. *The Technique*

Questioned document examiners are called on to perform a variety of tasks such as determining the sequence of strokes on a page and whether a particular ink formulation existed on the purported date of a writing.¹⁷⁷ However, the most common task performed is signature authentication—that is, deciding whether to attribute the handwriting on a document to a particular person. Here, the examiner compares known samples of the person’s writing to the questioned

173. FBI Review, *supra* note 158, at 7.

174. *Patterson*, 840 N.E.2d at 18 (“[T]he examiner apparently may take into account the distance separating the latent impressions, the orientation of the impressions, the pressure used to make the impression, and any other facts the examiner deems relevant. The record does not, however, indicate that there is any approved standardized method for making the determination that two or more print impressions have been made simultaneously.”).

175. The FBI review addressed this subject: “[I]f an item could only be held in a certain manner, then the only way of explaining the evidence is that the multiple prints are from the single person. In some cases, identifying simultaneous prints may infer, for example, the manner in which a knife was held.” FBI Review, *supra* note 158, at 8. However, the review found that there was not agreement on what constitutes a “simultaneous impression,” and therefore, more explicit guidelines were needed.

176. Irby Todd, *Do Experts Frequently Disagree?* 18 J. Forensic Sci. 455, 457–59 (1973).

177. Questioned document examinations cover a wide range of analyses: handwriting, hand printing, typewriting, mechanical impressions, altered documents, obliterated writing, indented writing, and charred documents. See 2 Paul C. Giannelli & Edward J. Imwinkelried, *Scientific Evidence* ch. 21 (4th ed. 2007).

document. In performing this comparison, examiners consider (1) class and (2) individual characteristics. Of class characteristics, two types are weighed: system¹⁷⁸ and group. People exhibiting system characteristics would include, for example, those who learned the Palmer method of cursive writing, taught in many schools. Such people should manifest some of the characteristics of that writing style. An example of people exhibiting group characteristics would include persons of certain nationalities who tend to have some writing mannerisms in common.¹⁷⁹ The writing of arthritic or blind persons also tends to exhibit some common general characteristics.¹⁸⁰

Individual characteristics take several forms: (1) the manner in which the author begins or ends the word, (2) the height of the letters, (3) the slant of the letters, (4) the shading of the letters, and (5) the distance between the words. An identification rarely rests on a single characteristic. More commonly, a combination of characteristics is the basis for an identification. As in fingerprint analysis, there is no universally accepted number of points of similarity required for an individuation opinion. As with fingerprints, the examiner's ultimate judgment is subjective.

There is one major difference, though, between the approaches taken by fingerprint analysts and questioned document examiners. As previously stated, the typical fingerprint analyst will give one of only three opinions: (1) the prints are unsuitable for analysis, (2) the suspect is definitely excluded, or (3) the latent print is definitely that of the suspect. In contrast, questioned document examiners recognize a wider range of permissible opinions: (1) definite identification, (2) strong probability of identification, (3) probable identification, (4) indication of identification, (5) no conclusion, (6) indication of nonauthorship, (7) probability of nonauthorship, (8) strong probability of nonauthorship, and (9) elimination.¹⁸¹ In short, in many cases, a questioned document examiner explicitly acknowledges the uncertainty of his or her opinion.¹⁸² Whether such a nine-level scale is justified is another matter.¹⁸³

178. See James A. Kelly, *Questioned Document Examination, in Scientific and Expert Evidence* 695, 698 (2d ed. 1981).

179. See Nellie Chang et al., *Investigation of Class Characteristics in English Handwriting of the Three Main Racial Groups: Chinese, Malay, and Indian in Singapore*, 50 J. Forensic Sci. 177 (2005); Robert J. Muehlberger, *Class Characteristics of Hispanic Writing in the Southeastern United States*, 34 J. Forensic Sci. 371 (1989); Sandra L. Ramsey, *The Cherokee Syllabary*, 39 J. Forensic Sci. 1039 (1994) (one of the landmark questioned document cases, *Hickory v. United States*, 151 U.S. 303 (1894), involved Cherokee writing); Marvin L. Simmer et al., *A Comparison of the Arabic Numerals One Through Nine, Written by Adults from Native English-Speaking vs. Non-Native English-Speaking Countries*, 15 J. Forensic Doc. Examination (2003).

180. See Larry S. Miller, *Forensic Examination of Arthritic Impaired Writings*, 15 J. Police Sci. & Admin. 51 (1987).

181. NRC Forensic Science Report, *supra* note 3, at 166.

182. See *id.* at 47.

183. See *United States v. Starzeczpyzel*, 880 F. Supp. 1027, 1048 (S.D.N.Y. 1995) ("No showing has been made, however, that FDEs can combine their first stage observations into such accurate conclusions as would justify a nine level scale.").

B. The Empirical Record

The 2009 NRC report included a section discussing questioned document examination. The report acknowledged that some tasks performed by examiners are similar in nature “to other forensic chemistry work.”¹⁸⁴ For example, some ink and paper analyses use the same hardware and rely on criteria as objective as many tests in forensic chemistry. In contrast, other analyses depend heavily on the examiner’s subjective judgment and do not have as “firm [a] scientific foundation” as the analysis of inks and paper.¹⁸⁵ In particular, the report focused on the typical task of deciding common authorship. With respect to that task, the report stated:

The scientific basis for handwriting comparisons needs to be strengthened. Recent studies have increased our understanding of the individuality and consistency of handwriting . . . and suggest that there may be a scientific basis for handwriting comparison, at least in the absence of intentional obfuscation or forgery. Although there has been only limited research to quantify the reliability and replicability of the practices used by trained document examiners, the committee agrees that there may be some value in handwriting analysis.¹⁸⁶

Until recently, the empirical record for signature authentication was sparse. Even today there are no population frequency studies establishing, for example, the incidence of persons who conclude their “w” with a certain lift. As a 1989 article commented,

our literature search for empirical evaluation of handwriting identification turned up one primitive and flawed validity study from nearly 50 years ago, one 1973 paper that raises the issue of consistency among examiners but presents only uncontrolled impressionistic and anecdotal information not qualifying as data in any rigorous sense, and a summary of one study in a 1978 government report. Beyond this, nothing.¹⁸⁷

This 1989 article then surveyed five proficiency tests administered by CTS in 1975, 1984, 1985, 1986, and 1987. The article set out the results from each of the tests¹⁸⁸ and then aggregated the data by computing the means for the various categories of answers: “A rather generous reading of the data would be that in 45% of the reports forensic document examiners reached the correct finding, in 36% they erred partially or completely, and in 19% they were unable to draw a conclusion.”¹⁸⁹

The above studies were conducted prior to *Daubert*, which was decided in 1993. After the first post-*Daubert* admissibility challenge to handwriting evidence

184. NRC Forensic Science Report, *supra* note 3, at 164.

185. *Id.* at 167.

186. *Id.* at 166–67.

187. Risinger et al., *supra* note 9, at 747.

188. *Id.* at 744 (1975 test), at 745 (1984 and 1985 tests), at 746 (1986 test), and at 747 (1987 test).

189. *Id.* at 747.

in 1995,¹⁹⁰ a number of research projects investigated two questions: (1) are experienced document examiners better at signature authentication than laypersons and (2) do experienced document examiners reach correct signature authentication decisions at a rate substantially above chance?

1. Comparison of experts and laypersons

Two Australian studies support the claim that experienced examiners are more competent at signature authentication tasks than laypersons. The first study was reported in 1999.¹⁹¹ In this study, document examiners chose the “inconclusive” option far more frequently than did the laypersons. However, in the cases in which a conclusion was reached, the overall error rate for lay subjects was 28%, compared with 2% for experts. More specifically, the lay error rate for false authentication was 7% while it was 0% for the experts. The second Australian study was released in 2002.¹⁹² Excluding “inconclusive” findings, the error rate for forensic document examiners was 5.8%; for laypersons, it was 23.5%.

In the United States, Dr. Moshe Kam, a computer scientist at Drexel University, has been the leading researcher in signature authentication. Dr. Kam and his colleagues have published five articles reporting experiments comparing the signature authentication expertise of document examiners and laypersons. Although the last study involved printing,¹⁹³ the initial four were related to cursive writing. In the first, excluding inconclusive findings, document examiners were correct 92.41% of the time and committed false elimination errors in 7.59% of their decisions.¹⁹⁴ Lay subjects were correct 72.84% of the time and made false elimination errors in 27.16% of their decisions. In the second through fourth studies, the researchers provided the laypersons with incentives, usually monetary, for correct decisions. In the fourth study, forgeries were called genuine only 0.5% of the time by experts but 6.5% of the time by laypersons.¹⁹⁵ Laypersons were 13 times more likely to err in concluding that a simulated document was genuine.

Some critics of Dr. Kam’s research have asserted that the tasks performed in the tests do not approximate the signature authentication challenges faced by

190. See *United States v. Starzeczyzel*, 880 F. Supp. 1027 (S.D.N.Y. 1995).

191. Bryan Found et al., *The Development of a Program for Characterizing Forensic Handwriting Examiners’ Expertise: Signature Examination Pilot Study*, 12 J. Forensic Doc. Examination 69, 72–76 (1999).

192. Jodi Sita et al., *Forensic Handwriting Examiners’ Expertise for Signature Comparison*, 47 J. Forensic Sci. 1117 (2002).

193. Moshe Kam et al., *Writer Identification Using Hand-Printed and Non-Hand-Printed Questioned Documents*, 48 J. Forensic Sci. 1 (2003).

194. Moshe Kam et al., *Proficiency of Professional Document Examiners in Writer Identification*, 39 J. Forensic Sci. 5 (1994).

195. Moshe Kam et al., *Signature Authentication by Forensic Document Examiners*, 46 J. Forensic Sci. 884 (2001); Moshe Kam et al., *The Effects of Monetary Incentives on Performance of Nonprofessionals in Document Examiners Proficiency Tests*, 43 J. Forensic Sci. 1000 (1998); Moshe Kam et al., *Writer Identification by Professional Document Examiners*, 42 J. Forensic Sci. 778 (1997).

examiners in real life.¹⁹⁶ In addition, critics have claimed that even the monetary incentives for the laypersons do not come close to equaling the powerful incentives that experts have to be careful in these tests.¹⁹⁷ Yet by now the empirical research record includes a substantial number of studies. With the exception of a 1975 German study,¹⁹⁸ the studies uniformly conclude that professional examiners are much more adept at signature authentication than laypersons.¹⁹⁹

2. Proficiency studies comparing experts' performance to chance

Numerous proficiency studies have been conducted in the United States²⁰⁰ and Australia.²⁰¹ Some of the American tests reported significant error rates. For example, on a 2001 test, excluding inconclusive findings, the false authentication rate was 22%, while the false elimination rate was 0%. Moreover, as previously stated, on the five CTS proficiency tests mentioned in the 1989 article, 36% of the participating examiners erred partially or completely.²⁰² Further, critics have claimed that some of the proficiency tests were far easier than the tasks encountered in actual practice,²⁰³ and that consequently, the studies tend to overstate examiners' proficiency.

196. D. Michael Risinger, *Cases Involving the Reliability of Handwriting Identification Expertise Since the Decision in Daubert*, 43 Tulsa L. Rev. 477, 490 (2007).

197. *Id.*

198. The German study included 25 experienced examiners, laypersons with no handwriting background, and some university students who had taken courses in handwriting psychology and comparison. On the one hand, the professional examiners outperformed the regular laypersons. The experts had a 14.7% error rate compared with the 34.4% rate for laypersons without any training. On the other hand, the university students had a lower aggregate error rate than the professional questioned document examiners. Wolfgang Conrad, *Empirische Untersuchungen über die Urteilsgute verschiedener Gruppen von Laien und Sachverständigen bei der Unterscheidung authentischer und gefälschter Unterschriften* [Empirical Studies Regarding the Quality of Assessments of Various Groups of Lay Persons and Experts in Differentiating Between Authentic and Forged Signatures], 156 *Archiv für Kriminologie* 169–83 (1975).

199. See Roger Park, *Signature Identification in the Light of Science and Experience*, 59 *Hastings L.J.* 1101, 1135–36 (2008).

200. *E.g.*, Collaborative Testing Service (CTS), Questioned Document Examination, Report No. 92-6 (1992); CTS, Questioned Document Examination, Report No. 9406 (1994); CTS, Questioned Document Examination, Report No. 9606 (1996); CTS, Forensic Testing Program, Handwriting Examination, Report No. 9714 (1997); CTS, Forensic Testing Program, Handwriting Examination, Report No. 9814 (1998); CTS, Forensic Testing Program, Handwriting Examination, Test No. 99-524 (1999); CTS, Forensic Testing Program, Handwriting Examination, Test No. 00-524 (2000); CTS, Forensic Testing Program, Handwriting Examination, Test No. 01-524 (2001); CTS, Forensic Testing Program, Handwriting Examination, Test No. 02-524 (2003); available at <http://www.ctsforensics.com/reports/main.aspx>.

201. Bryan Found & Doug Rogers, *The Probative Character of Forensic Handwriting Examiners' Identification and Elimination Opinions on Questioned Signatures*, 178 *Forensic Sci. Int'l* 54 (2008).

202. Risinger et al., *supra* note 9, at 747–48.

203. Risinger, *supra* note 196, at 485.

The CTS proficiency test results for the 1978–2005 period addressed the comparison of known and questioned signatures and other writings to determine authorship. In other exercises participants were asked to examine a variety of mechanical impressions on paper and the use of photocopying and inks.

- Between 1978 and 1999,²⁰⁴ fewer than 5% of the mechanical impression comparisons were in error, but 10% of the replies were inconclusive where the examiner should have excluded the impressions as having a common source. With regard to handwriting comparisons, the examiners did very well on the straightforward comparisons, with almost 100% of the comparisons correct. However, in more challenging tests, such as those involving multiple authors, as high as 25% of the replies were inconclusive and nearly 10% of the author associations were incorrect.
- In the 2000–2005 time period, the participants generally performed very well (some approaching 99% correct responses) in determining the genuineness of documents where text in a document had been manipulated or where documents had been altered with various pens and inks. The handwriting exercises were not as successful; in those exercises, comparisons of questioned and known writings were correct about 92% of the time, inconclusive 7% of the time, and incorrect 1% of the time. Nearly all incorrect responses occurred where participants reported handwriting to be of common origin when it was not.

During these tests, some examiners characterized the tests as too easy, while others described them as realistic and very challenging.

Thus, the results of the most recent proficiency studies are encouraging. Moreover, the data in the five proficiency tests discussed in the 1989 article²⁰⁵ can be subject to differing interpretation. The critics of questioned document examination sometimes suggest that the results of the 1985 test in particular prove that signature authentication has “a high error rate.”²⁰⁶ However,

[t]hese results can be characterized in different ways. [Another] way of viewing the result would be to disaggregate the specific decisions made by the experts. . . . [S]uppose that a teacher gives a multiple-choice test containing fifty questions. There are different ways that the results could be reported. One could calculate the percentage of students who got any of the fifty questions wrong, and report that as the error rate. A more customary approach would be to treat

204. John I. Thornton & Joseph L. Peterson, *The General Assumptions and Rationale of Forensic Identification*, in 4 *Modern Scientific Evidence: The Law and Science of Expert Testimony*, *supra* note 104, § 29:40, at 54.

205. Risinger et al., *supra* note 9.

206. Park, *supra* note 199, at 1113.

each question as a separate task, and report the error rate as the mean percentage of questions answered incorrectly.²⁰⁷

If the specific decisions made by the examiners were disaggregated, each examiner had to make 66 decisions regarding whether certain pairs of signatures were written by the same person.²⁰⁸ Under this approach, the false authentication error rate was 3.8%, and the false elimination error rate was 4.5%.²⁰⁹ In that light, even the 1985 study supports the contention that examiners perform signature authentication tasks at a validity rate considerably exceeding chance.

C. Case Law Development

Although the nineteenth-century cases were skeptical of handwriting expertise,²¹⁰ in the twentieth century the testimony in leading cases, such as the Lindbergh prosecution, helped the discipline gain judicial acceptance. There was little dispute that handwriting comparison testimony was admissible at the time the Federal Rules of Evidence were enacted in 1975. Rule 901(b)(3) recognized that a document could be authenticated by an expert, and the drafters explicitly mentioned handwriting comparison “testimony of expert witnesses.”²¹¹

The first significant admissibility challenge under *Daubert* was mounted in *United States v. Starzecpyzel*.²¹² In that case, the district court concluded that “forensic document examination, despite the existence of a certification program, professional journals and other trappings of science, cannot, after *Daubert*, be regarded as ‘scientific . . . knowledge.’”²¹³ Nonetheless, the court did not exclude handwriting comparison testimony. Instead, the court admitted the individuation testimony as nonscientific “technical” evidence.²¹⁴ *Starzecpyzel* prompted more attacks that questioned the lack of empirical validation in the field.²¹⁵

207. *Id.* at 1114.

208. *Id.* at 1115.

209. *Id.* at 1116.

210. See *Strother v. Lucas*, 31 U.S. 763, 767 (1832); *Phoenix Fire Ins. Co. v. Philip*, 13 Wend. 81, 82–84 (N.Y. Sup. Ct. 1834).

211. Fed. R. Evid. 901(b)(3) advisory committee’s note.

212. 880 F. Supp. 1027 (S.D.N.Y. 1995).

213. *Id.* at 1038.

214. *Kumho Tire* later called this aspect of the *Starzecpyzel* opinion into question because *Kumho* held that the reliability requirement applies to all types of expertise—“scientific,” “technical,” or “specialized.” Moreover, the Supreme Court indicated that the *Daubert* factors, including empirical testing, may be applicable to technical expertise. Some aspects of handwriting can and have been tested.

215. See, e.g., *United States v. Hidalgo*, 229 F. Supp. 2d 961, 967 (D. Ariz. 2002) (“Because the principle of uniqueness is without empirical support, we conclude that a document examiner will not be permitted to testify that the maker of a known document is the maker of the questioned document. Nor will a document examiner be able to testify as to identity in terms of probabilities.”).

As of the date of this publication, there is a three-way split of authority. The majority of courts permit examiners to express individuation opinions.²¹⁶ As one court noted, “all six circuits that have addressed the admissibility of handwriting expert [testimony] . . . [have] determined that it can satisfy the reliability threshold” for nonscientific expertise.²¹⁷ In contrast, several courts have excluded expert testimony,²¹⁸ although one involved handprinting²¹⁹ and another Japanese handprinting.²²⁰ Many district courts have endorsed a third view. These courts limit the reach of the examiner’s opinion, permitting expert testimony about similarities and dissimilarities between exemplars but not an ultimate conclusion that the defendant was the author (“common authorship” opinion) of the questioned document.²²¹ The expert is allowed to testify about “the specific similarities and idiosyncrasies between the known writings and the questioned writings, as well as testimony regarding, for example, how frequently or infrequently in his experience, [the expert] has seen a particular idiosyncrasy.”²²² As the justification for this limitation, these courts often state that the examiners’ claimed ability to individuate lacks “empirical support.”²²³

216. See, e.g., *United States v. Prime*, 363 F.3d 1028, 1033 (9th Cir. 2004); *United States v. Crisp*, 324 F.3d 261, 265–71 (4th Cir. 2003); *United States v. Jolivet*, 224 F.3d 902, 906 (8th Cir. 2000) (affirming the introduction of expert testimony that it was likely that the accused wrote the questioned documents); *United States v. Velasquez*, 64 F.3d 844, 848–52 (3d Cir. 1995); *United States v. Ruth*, 42 M.J. 730, 732 (A. Ct. Crim. App. 1995), *aff’d on other grounds*, 46 M.J. 1 (C.A.A.F. 1997); *United States v. Morris*, No. 06–87–DCR, 2006 U.S. Dist. LEXIS 53983, *5 (E.D. Ky. July 20, 2006); *Orix Fin. Servs. v. Thunder Ridge Energy, Inc.*, No. 01Civ. 4788 (RJH) (HBP). 2005 U.S. Dist. LEXIS 41889 (S.D.N.Y. Dec. 29, 2005).

217. *Prime*, 363 F.3d at 1034.

218. *United States v. Lewis*, 220 F. Supp. 2d 548 (S.D. W. Va. 2002).

219. *United States v. Saelee*, 162 F. Supp. 2d 1097 (D. Alaska 2001).

220. *United States v. Fujii*, 152 F. Supp. 2d 939, 940 (N.D. Ill. 2000) (holding expert testimony concerning Japanese handprinting inadmissible: “Handwriting analysis does not stand up well under the *Daubert* standards. Despite its long history of use and acceptance, validation studies supporting its reliability are few, and the few that exist have been criticized for methodological flaws.”).

221. See, e.g., *United States v. Oskowitz*, 294 F. Supp. 2d 379, 384 (E.D.N.Y. 2003) (“Many other district courts have similarly permitted a handwriting expert to analyze a writing sample for the jury without permitting the expert to offer an opinion on the ultimate question of authorship.”); *United States v. Rutherford*, 104 F. Supp. 2d 1190, 1194 (D. Neb. 2000) (“[T]he Court concludes that FDE Rauscher’s testimony meets the requirements of Rule 702 to the extent that he limits his testimony to identifying and explaining the similarities and dissimilarities between the known exemplars and the questioned documents. FDE Rauscher is precluded from rendering any ultimate conclusions on authorship of the questioned documents and is similarly precluded from testifying to the degree of confidence or certainty on which his opinions are based.”); *United States v. Hines*, 55 F. Supp. 2d 62, 69 (D. Mass. 1999) (expert testimony concerning the general similarities and differences between a defendant’s handwriting exemplar and a stick-up note was admissible while the specific conclusion that the defendant was the author was not).

222. *United States v. Van Wyk*, 83 F. Supp. 2d 515, 524 (D.N.J. 2000).

223. *United States v. Hidalgo*, 229 F. Supp. 2d 961, 967 (D. Ariz. 2002).

VIII. Firearms Identification Evidence

It is widely considered that the first written reference to firearms identification (popularly known as “ballistics”) in the United States appeared in 1900.²²⁴ In the 1920s, the technique gained considerable attention because of the work of Calvin Goddard²²⁵ and played a controversial role in the Sacco and Vanzetti case during the same decade.²²⁶ Goddard also analyzed the bullet evidence in the St. Valentine’s Day Massacre in 1929, in which five gangsters and two acquaintances were gunned down in Chicago.²²⁷ In 1923, the Illinois Supreme Court wrote that positive identification of a bullet was not only impossible but “preposterous.”²²⁸ Seven years later, however, that court did an about-face and became one of the first courts in this country to admit firearms identification evidence.²²⁹ The technique subsequently gained widespread judicial acceptance and was not seriously challenged until recently.

A. The Technique

1. Firearms

Typically, three types of firearms—rifles, handguns, and shotguns—are encountered in criminal investigations.²³⁰ The barrels of modern rifles and handguns are *rifled*; that is, parallel spiral grooves are cut into the inner surface (bore) of the barrel. The surfaces between the grooves are called *lands*. The lands and grooves twist in a direction: right twist or left twist. For each type of firearm produced, the manufacturer specifies the number of lands and grooves, the direction of twist, the angle of twist (pitch), the depth of the grooves, and the width of the lands and grooves. As a bullet passes through the bore, the lands and grooves force the

224. See Albert Llewellyn Hall, *The Missile and the Weapon*, 39 Buff. Med. J. 727 (1900).

225. Calvin Goddard, often credited as the “father” of firearms identification, was responsible for much of the early work on the subject. E.g., Calvin Goddard, *Scientific Identification of Firearms and Bullets*, 17 J. Crim. L., Criminology & Police Sci. 254 (1926).

226. See Joughin & Morgan, *supra* note 8, at 15 (The firearms identification testimony was “carelessly assembled, incompletely and confusedly presented, and . . . beyond the comprehension” of the jury); Starrs, *supra* note 8, at 630 (Part I), 1050 (Part II).

227. See Calvin Goddard, *The Valentine Day Massacre: A Study in Ammunition-Tracing*, 1 Am. J. Police Sci. 60, 76 (1930) (“Since two of the members of the execution squad had worn police uniforms, and since it had been subsequently intimated by various persons that the wearers of the uniforms might really have been policeman rather than disguised gangsters, it became a matter of no little importance to ascertain, if possible, whether these rumors had any foundation in fact.”); Jim Ritter, *St. Valentine’s Hit Spurred Creation of Nation’s First Lab*, Chicago Sun-Times, Feb. 9, 1997, at 40 (“Sixty-eight years ago this Friday, Al Capone’s hit men, dressed as cops, gunned down seven men in the Clark Street headquarters of rival mobster Bugs Moran.”).

228. *People v. Berkman*, 139 N.E. 91, 94 (Ill. 1923).

229. *People v. Fisher*, 172 N.E. 743, 754 (Ill. 1930).

230. Other types of firearms, such as machine guns, tear gas guns, zip guns, and flare guns, may also be examined.

bullet to rotate, giving it stability in flight and thus increased accuracy. Shotguns are smooth-bore firearms; they do not have lands and grooves.

Rifles and handguns are classified according to their caliber. The caliber is the diameter of the bore of the firearm; the caliber is expressed in either hundredths or thousandths of an inch (e.g., .22, .45, .357 caliber) or millimeters (e.g., 7.62 mm).²³¹ The two major types of handguns are revolvers²³² and semiautomatic pistols. A major difference between the two is that when a semiautomatic pistol is fired, the cartridge case is automatically ejected and, if recovered at the crime scene, could help link the case to the firearm from which it was fired. In contrast, when a revolver is discharged the case is not ejected.

2. Ammunition

Rifle and handgun cartridges consist of the projectile (bullet),²³³ case,²³⁴ propellant (powder), and primer. The primer contains a small amount of an explosive mixture, which detonates when struck by the firing pin. When the firing pin detonates the primer, an explosion occurs that ignites the propellant. The most common modern propellant is smokeless powder.

3. Class characteristics

Firearms identifications may be based on either bullet or cartridge case examinations. Identifying features include class, subclass, and individual characteristics.

The class characteristics of a firearm result from design factors and are determined prior to manufacture. They include the following caliber and rifling specifications: (1) the land and groove diameters, (2) the direction of rifling (left or right twist), (3) the number of lands and grooves, (4) the width of the lands and grooves, and (5) the degree of the rifling twist.²³⁵ Generally, a .38-caliber bullet with six land and groove impressions and with a right twist could have been fired only from a firearm with these same characteristics. Such a bullet could not have been fired from a .32-caliber firearm, or from a .38-caliber firearm with a different number of lands and grooves or a left twist. In sum, if the class characteristics do not match, the firearm could not have fired the bullet and is excluded.

231. The caliber is measured from land to land in a rifled weapon. Typically, the designated caliber is more an approximation than an accurate measurement. See 1 J. Howard Mathews, *Firearms Identification* 17 (1962) (“‘nominal caliber’ would be a more proper term”).

232. Revolvers have a cylindrical magazine that rotates behind the barrel. The cylinder typically holds five to nine cartridges, each within a separate chamber. When a revolver is fired, the cylinder rotates and the next chamber is aligned with the barrel. A single-action revolver requires the manual cocking of the hammer; in a double-action revolver the trigger cocks the hammer.

233. Bullets are generally composed of lead and small amounts of other elements (hardeners). They may be completely covered (jacketed) with another metal or partially covered (semijacketed).

234. Cartridge cases are generally made of brass.

235. 1 Mathews, *supra* note 231, at 17.

4. Subclass characteristics

Subclass characteristics are produced at the time of manufacture and are shared by a discrete subset of weapons in a production run or “batch.” According to the Association of Firearm and Tool Mark Examiners (AFTE),²³⁶ subclass characteristics are discernible surface features that are more restrictive than class characteristics in that they are (1) “produced incidental to manufacture,” (2) “relate to a smaller group source (a subset to which they belong),” and (3) can arise from a source that changes over time.²³⁷ The AFTE states that “[c]aution should be exercised in distinguishing subclass characteristics from class characteristics.”²³⁸

5. Individual characteristics

Bullet identification involves a comparison of the evidence bullet and a test bullet fired from the firearm.²³⁹ The two bullets are examined by means of a comparison microscope, which permits a split-screen view of the two bullets and manipulation in order to attempt to align the striations (marks) on the two bullets.

Barrels are machined during the manufacturing process, and imperfections in the tools used in the machining process are imprinted on the bore.²⁴⁰ The subsequent use of the firearm adds further individual imperfections. For example, mechanical action (erosion) caused by the friction of bullets passing through the bore of the firearm produces accidental imperfections. Similarly, chemical action (corrosion) caused by moisture (rust), as well as primer and propellant chemicals, produce other imperfections.

When a bullet is fired, microscopic striations are imprinted on the bullet surface as it passes through the bore of the firearm. These bullet markings are produced by the imperfections in the bore. Because these imperfections are randomly produced, examiners assume that they are unique to each firearm.²⁴¹ Although the assumption is plausible, there is no statistical basis for this assumption.²⁴²

236. AFTE is the leading professional organization in the field. There is also the Scientific Working Group for Firearms and Toolmarks (SWGUN), which promulgates guidelines for examiners.

237. *Theory of Identification, Association of Firearm and Toolmark Examiners*, 30 AFTE J. 86, 88 (1998) [hereinafter *AFTE Theory*].

238. *Id.*

239. Test bullets are obtained by firing a firearm into a recovery box or bullet trap, which is usually filled with cotton, or into a recovery tank, which is filled with water.

240. “No two barrels are microscopically identical, as the surfaces of their bores all possess individual and characteristic markings.” Gerald Burrard, *The Identification of Firearms and Forensic Ballistics* 138 (1962).

241. 1 Mathews, *supra* note 231, at 3 (“Experience has shown that no two firearms, even those of the same make and model and made consecutively by the same tools, will produce the same markings on a bullet or a cartridge.”).

242. Alfred A. Biasotti, *The Principles of Evidence Evaluation as Applied to Firearms and Tool Mark Identification*, 9 J. Forensic Sci. 428, 432 (1964) (“[W]e lack the fundamental statistical data needed to

Although an identification is based on objective data (the striations on the bullet surface), the AFTE explains that the examiner's individuation is essentially a subjective judgment. The AFTE describes the traditional pattern recognition methodology as "subjective in nature, founded on scientific principles and based on the examiner's training and experience."²⁴³ There are no objective criteria governing this determination: "Ultimately, unless other issues are involved, it remains for the examiner to determine for himself the modicum of proof necessary to arrive at a definitive opinion."²⁴⁴

The condition of a firearm or evidence bullet may preclude an identification. For example, there may be insufficient marks on the bullet or, because of mutilation, an insufficient amount of the bullet may have been recovered. Likewise, if the bore of the firearm has changed significantly as a result of erosion or corrosion, an identification may be impossible. (Unlike fingerprints, firearms change over time.) In these situations, the examiner may render a "no conclusion" determination. Such a conclusion, however, may have some evidentiary value even if the examiner cannot form an individuation opinion; that is, the firearm could have fired the bullet if the class characteristics match.

6. Consecutive matching striae

In an attempt to make firearms identification more objective, some commentators advocate a technique known as consecutive matching striae (CMS). As the name implies, this method is based on finding a specified number of consecutive matching striae on two bullets. Other commentators have questioned this approach,²⁴⁵ and it remains a minority position.²⁴⁶

7. Cartridge identification

Cartridge case identification is based on the same theory of random markings as bullet identification.²⁴⁷ As with barrels, defects produced in the manufacturing

develop verifiable criteria."); see also Alfred A. Biasotti, *A Statistical Study of the Individual Characteristics of Fired Bullets*, 4 J. Forensic Sci. 34 (1959).

243. *AFTE Theory*, *supra* note 237, at 86.

244. Laboratory Proficiency Test, *supra* note 81, at 207; see also Alfred A. Biasotti, *The Principles of Evidence Evaluation as Applied to Firearms and Tool Mark Identification*, *supra* note 242, at 429 ("In general, the texts on firearms identification take the position that each practitioner must develop his own intuitive criteria of identity gained through practical experience.").

245. See Stephen G. Bunch, *Consecutive Matching Striation Criteria: A General Critique*, 45 J. Forensic Sci. 955, 955 (2000) (finding the traditional methodology superior: "[P]resent-day firearm identification, in the final analysis is subjective.").

246. Roger C. Nichols, *Firearm and Toolmark Identification Criteria: A Review of the Literature, Part II*, 48 J. Forensic Sci. 318, 326 (2003) (CMS "has not been promoted as an alternative [to traditional pattern recognition], but as a numerical threshold.").

247. Burrard, *supra* note 240, at 107. However, bullet and cartridge case identifications differ in several respects. Because the bullet is traveling through the barrel at the time it is imprinted with the

process leave distinctive characteristics on the breech face, firing pin, chamber, extractor, and ejector. Subsequent use of the firearm produces additional defects. When the trigger is pulled, the firing pin strikes the primer of the cartridge, causing the primer to detonate. This detonation ignites the propellant (powder). In the process of combustion, the powder is converted rapidly into gases. The pressure produced by this process propels the bullet from the weapon and also forces the base of the cartridge case backward against the breech face, imprinting breech face marks on the base of the cartridge case. Similarly, the firing pin, ejector, and extractor may leave characteristic marks on a cartridge case.²⁴⁸

Cartridge case identification involves a comparison of the cartridge case recovered at the crime scene and a test cartridge case obtained from the firearm after it has been fired. Shotgun shell casings may be identified in this way, as well. As in bullet identification, the comparison microscope is used in the examination. According to AFTE, “interpretation of toolmark individualization and identification is still considered to be subjective in nature, based on one’s training and experience.”²⁴⁹

8. Automated identification systems

“These ballistic imaging systems use the powerful searching capabilities of the computer to match the images of recovered crime scene evidence against digitized images stored in a computer database.”²⁵⁰ The current system is the Integrated Ballistics Information System (IBIS).²⁵¹ Automated systems “give[] firearms examiners the ability to screen virtually unlimited numbers of bullets and cartridge casings for possible matches.”²⁵² These systems identify a number of candidate matches. They do not replace the examiner, who still must make the final comparison: “‘High Confidence’ candidates (likely hits) are referred to a firearms examiner for examination on a comparison microscope.”²⁵³ The examiner need

bore imperfections, these marks are “sliding” imprints, called striated marks. In contrast, the cartridge case receives “static” imprints, called impressed marks. *Id.* at 145.

248. Ejector and extractor marks by themselves may indicate only that the cartridge case had been *loaded in*, not fired from, a particular firearm.

249. Eliot Springer, *Toolmark Examinations—A Review of Its Development in the Literature*, 40 *J. Forensic Sci.* 964, 966–67 (1995).

250. *Benchmark Evaluation Studies of the Bulletproof and Drugfire Ballistic Imaging Systems*, 22 *Crime Lab. Digest* 51 (1995); see also Jan De Kinder & Monica Bonfanti, *Automated Comparisons of Bullet Striations Based on 3D Topography*, 101 *Forensic Sci. Int’l* 85, 86 (1999) (“[A]n automatic system will cut the time demanding and tedious manual searches for one specific item in large open case files.”).

251. See Jan De Kinder et al., *Reference Ballistic Imaging Database Performance*, 140 *Forensic Sci. Int’l* 207 (2004); Ruprecht Nennstiel & Joachim Rahm, *An Experience Report Regarding the Performance of the IBIS™ Correlator*, 51 *J. Forensic Sci.* 24 (2006).

252. Richard E. Tontarski & Robert M. Thompson, *Automated Firearms Evidence Comparison: A Forensic Tool for Firearms Identification—An Update*, 43 *J. Forensic Sci.* 641, 641 (1998).

253. *Id.*

not accept the highest ranked candidate identified by the system. For that matter, the examiner may reject all the candidates.

9. Toolmarks

Toolmark identifications rest on essentially the same theory as firearms identifications.²⁵⁴ Tools have both (1) class characteristics and (2) individual characteristics; the latter are accidental imperfections produced by the machining process and subsequent use. When the tool is used, these characteristics are sometimes imparted onto the surface of another object struck by the tool. Toolmarks may be impressions (compression marks), striations (friction or scrape marks), or a combination of both.²⁵⁵ Fracture matches constitute another type of examination.

The marks may be left on a variety of different materials, such as wood or metal. In some cases, only class characteristics can be matched. For example, it may be possible to identify a mark (impression) left on a piece of wood as having been produced by a hammer, punch, or screwdriver. A comparison of the mark and the evidence tool may establish the size of the tool (another class characteristic). Unusual features of the tool, such as a chip, may permit a positive identification. Striations caused by scraping with a tool can also produce distinguishing marks in much the same way that striations are imprinted on a bullet when a firearm is discharged. This type of examination has the same limitations as firearms identification: “[T]he characteristics of a tool will change with use.”²⁵⁶

Firearms identification could be considered a subspecialty of toolmark identification; the firearm (tool) imprints its individual characteristics on the bullet. However, the markings on a bullet or cartridge case are imprinted in roughly the same way every time a firearm is fired. In contrast, toolmark analysis can be more complicated because a tool can be employed in a variety of different ways, each producing a different mark: “[I]n toolmark work the angle at which the tool was used must be duplicated in the test standard, pressures must be dealt with, and the degree of hardness of metals and other materials must be taken into account.”²⁵⁷

The comparison microscope is also used in this examination. As with firearms identification testimony, toolmark identification testimony is based on the subjective judgment of the examiner, who determines whether sufficient marks of

254. See Biasotti, *The Principles of Evidence Evaluation as Applied to Firearms and Tool Mark Identification*, *supra* note 242; see also Springer, *supra* note 249, at 964 (“The identification is based . . . on a series of scratches, depressions, and other marks which the tool leaves on the object it comes into contact with. The combination of these various marks ha[s] been termed toolmarks and the claim is that every instrument can impart a mark individual to itself.”).

255. David Q. Burd & Roger S. Greene, *Tool Mark Examination Techniques*, 2 J. Forensic Sci. 297, 298 (1957).

256. Emmett M. Flynn, *Toolmark Identification*, 2 J. Forensic Sci. 95, 102 (1957).

257. *Id.* at 105.

similarity are present to permit an identification.²⁵⁸ There are no objective criteria governing the determination of whether there is a match.²⁵⁹

B. *The Empirical Record*

In its 2009 report, NRC summarized the state of the research as follows:

Because not enough is known about the variabilities among individual tools and guns, we are not able to specify how many points of similarity are necessary for a given level of confidence in the result. Sufficient studies have not been done to understand the reliability and repeatability of the methods. The committee agrees that class characteristics are helpful in narrowing the pool of tools that may have left a distinctive mark. Individual patterns from manufacture or from wear might, in some cases, be distinctive enough to suggest one particular source, but additional studies should be performed to make the process of individualization more precise and repeatable.²⁶⁰

The 1978 Crime Laboratory Proficiency Testing Program reported mixed results on firearms identification tests. In one test, 5.3% of the participating laboratories misidentified firearms evidence, and in another test 13.6% erred. These tests involved bullet and cartridge case comparisons. The Project Advisory Committee considered these errors “particularly grave in nature” and concluded that they probably resulted from carelessness, inexperience, or inadequate supervision.²⁶¹ A third test required the examination of two bullets and two cartridge cases to identify the “most probable weapon” from which each was fired. The error rate was 28.2%.

In later tests,

[e]xaminers generally did very well in making the comparisons. For all fifteen tests combined, examiners made a total of 2106 [bullet and cartridge case] comparisons and provided responses which agreed with the manufacturer responses 88% of the time, disagreed in only 1.4% of responses, and reported inconclusive results in 10% of cases.²⁶²

258. See Springer, *supra* note 249, at 966–67 (“According to the Association of Firearms and Toolmarks Examiners’ Criteria for Identification Committee, interpretation of toolmark individualization and identification is still considered to be subjective in nature, based on one’s training and experience.”).

259. As one commentator has noted: “[I]t is not possible at present to categorically state the number and percentage of the [striation] lines which must correspond.” Burd & Greene, *supra* note 255, at 310.

260. *Id.*

261. Laboratory Proficiency Test, *supra* note 81, at 207–08.

262. Peterson & Markham, *supra* note 82, at 1018. The authors also stated:

The performance of laboratories in the firearms tests was comparable to that of the earlier LEAA study, although the rate of successful identifications actually was slightly over—88% vs. 91%. Laboratories cut the rate of errant identifications by half (3% to 1.4%) but the rate of inconclusive responses doubled, from 5% to 10%.

Id. at 1019.

Proficiency testing on toolmark examinations has also been reported.²⁶³

For the period 1978–1999, firearms examiners performed well on their CTS proficiency tests, with only 2% to 3% of their comparisons incorrect, but with 10% to 13% of their responses inconclusive.²⁶⁴ The scenarios that accompanied the test materials asked examiners to compare test-fired bullets and/or cartridge cases with evidence projectiles found at a crime scene. Between 2000 and 2005, participants, again, performed very well, averaging less than 1% incorrect responses, but with inconclusive results about 10% of the time. Most of the inconclusive results in these tests occurred where bullets and/or cartridge cases were actually fired from different weapons. Examiners frequently stated they were unable to reach the proper conclusion because they did not have the actual weapon with which they could perform their own test fires of ammunition.

In CTS toolmark proficiency comparisons, laboratories were asked to compare marks made with such tools as screwdrivers, bolt cutters, hammers, and hand-stamps. In some cases, tools were supplied to participants, but in most cases they were given only test marks. Over the entire 1978–2005 period, fewer than 5% of responses were in error, but individual test results varied substantially. In some cases, 30% to 40% of replies were inconclusive, because laboratories were unsure if the blade of the tool in question might have been altered between the time(s) different markings had been made. During the final 6-year period reviewed (2000–2005), laboratories averaged a 1% incorrect comparison rate for toolmarks. Inconclusive responses remained high (30% and greater) and, together with firearms testing, constitute the evidence category where evidence comparisons have the highest rates of inconclusive responses.

Questions have arisen concerning the significance of these tests. First, such testing is not required of all firearms examiners, only those working in laboratories voluntarily seeking accreditation by the ASCLD. In short, “the sample is self-selecting and may not be representative of the complete universe of firearms examiners.”²⁶⁵ Second, the examinations are not blind—that is, examiners know when they are being tested. Thus, the examiner may be more meticulous and careful than in ordinary case work. Third, the results of an evaluation can vary, depending on whether an “inconclusive” answer is counted. Fourth, the rigor of the examinations has been questioned. According to one witness, in a 2005 test involving cartridge case comparisons, *none* of the 255 test-takers nationwide answered incorrectly. The court observed: “One could read these results to mean that the technique is foolproof, but the results might instead indicate that the test was somewhat elementary.”²⁶⁶

263. *Id.* at 1025 (“Overall, laboratories performed not as well on the toolmark tests as they did on the firearms tests.”).

264. Thornton & Peterson, *supra* note 204, § 29:47, at 66.

265. United States v. Monteiro, 407 F. Supp. 2d 351, 367 (D. Mass. 2006).

266. *Id.*

In 2008, NAS published a report on computer imaging of bullets.²⁶⁷ Although firearms identification was not the primary focus of the investigation, a section of the report commented on this subject.²⁶⁸ After surveying the literature on the uniqueness, reproducibility, and permanence of individual characteristics, the committee noted that “[m]ost of these studies are limited in scale and have been conducted by firearms examiners (and examiners in training) in state and local law enforcement laboratories as adjuncts to their regular casework.”²⁶⁹ The report concluded: “The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.”²⁷⁰ This statement, however, was qualified:

There is one baseline level of credibility . . . that must be demonstrated lest any discussion of ballistic imaging be rendered moot—namely, that there is at least some “signal” that may be detected. In other words, the creation of toolmarks must not be so random and volatile that there is no reason to believe that any similar and matchable marks exist on two exhibits fired from the same gun. The existing research, and the field’s general acceptance in legal proceedings for several decades, is more than adequate testimony to that baseline level. Beyond that level, we neither endorse nor oppose the fundamental assumptions. Our review in this chapter is not—and is not meant to be—a full weighing of evidence for or against the assumptions, but it is ample enough to suggest that they are not fully settled, mechanically or empirically.

Another point follows directly: *Additional general research on the uniqueness and reproducibility of firearms-related toolmarks would have to be done if the basic premises of firearms identification are to be put on a more solid scientific footing.*²⁷¹

The 2008 report cautioned:

Conclusions drawn in firearms identification should not be made to imply the presence of a firm statistical basis when none has been demonstrated. Specifically, . . . examiners tend to cast their assessments in bold absolutes, commonly asserting that a match can be made “to the exclusion of all other firearms in the world.” Such comments cloak an inherently subjective assessment of a match with an extreme probability statement that has no firm grounding and unrealistically implies an error rate of zero.²⁷²

267. National Research Council, *Ballistic Imaging* (2008), available at http://www.nap.edu/catalog.php?record_id=12162.

268. The committee was asked to assess the feasibility, accuracy, reliability, and technical capability of developing and using a national ballistic database as an aid to criminal investigations. It concluded: (1) “A national reference ballistic image database of all new and imported guns is not advisable at this time.” (2) “NIBIN can and should be made more effective through operational and technological improvements.” *Id.* at 5.

269. *Id.* at 70.

270. *Id.* at 81.

271. *Id.* at 81–82.

272. *Id.* at 82.

The issue of the adequacy of the empirical basis of firearms identification expertise remains in dispute,²⁷³ and research is ongoing. A recent study reported testing concerning 10 consecutively rifled Ruger pistol barrels. In 463 tests during the study, no false positives were reported; 8 inconclusive results were reported.²⁷⁴ “But the capsule summaries [in this study] suggest a heavy reliance on the subjective findings of examiners rather than on the rigorous quantification and analysis of sources of variability.”²⁷⁵

C. Case Law Development

Firearms identification developed in the early part of the last century, and by 1930, courts were admitting evidence based on this technique.²⁷⁶ Subsequent cases followed these precedents, admitting evidence of bullet,²⁷⁷ cartridge case,²⁷⁸ and shot shell²⁷⁹ identifications. A number of courts have also permitted an expert to

273. Compare Roger G. Nichols, *Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges*, 52 J. Forensic Sci. 586 (2007), with Adina Schwartz, *Commentary on “Nichols, R.G., Defending the scientific foundations of the firearms and tool mark identification discipline: Responding to recent challenges*, J. Forensic Sci. 52(3):586-94 (2007),”

52 J. Forensic Sci. 1414 (2007) (responding to Nichols). Moreover, AFTE disputed the Academy’s conclusions. See *The Response of the Association of Firearm and Tool Mark Examiners to the National Academy of Sciences 2008 Report Assessing the Feasibility, Accuracy, and Capability of a National Ballistic Database August 20, 2008*, 40 AFTE J. 234 (2008) (concluding that underlying assumptions of uniqueness and reproducibility have been demonstrated, and the implication that there is no statistical basis is unwarranted); see also Adina Schwartz, *A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification*, 6 Colum. Sci. & Tech. L. Rev. 2 (2005).

274. James E. Hamby et al., *The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels—A Research Project Involving 468 Participants from 19 Countries*, 41 AFTE J. 99 (Spring 2009).

275. NRC Forensic Science Report, *supra* note 3, at 155.

276. *E.g.*, *People v. Fisher*, 172 N.E. 743 (Ill. 1930); *Evans v. Commonwealth*, 19 S.W.2d 1091 (Ky. 1929); *Burchett v. State*, 172 N.E. 555 (Ohio Ct. App. 1930).

277. *E.g.*, *United States v. Wolff*, 5 M.J. 923, 926 (N.C.M.R. 1978); *State v. Mack*, 653 N.E.2d 329, 337 (Ohio 1995) (The examiner “compared the test shot with the morgue bullet recovered from the victim, . . . and the spent shell casings recovered from the crime scene, concluding that all had been discharged from appellant’s gun.”).

278. *E.g.*, *Bentley v. Scully*, 41 F.3d 818, 825 (2d Cir. 1994) (“[A] ballistic expert found that the spent nine millimeter bullet casing recovered from the scene of the shooting was fired from the pistol found on the rooftop.”); *State v. Samonte*, 928 P.2d 1, 6 (Haw. 1996) (“Upon examining the striation patterns on the casings, [the examiner] concluded that the casing she had fired matched six casings that police had recovered from the house.”).

279. *E.g.*, *Williams v. State*, 384 So. 2d 1205, 1210–11 (Ala. Crim. App. 1980); *Burge v. State*, 282 So. 2d 223, 229 (Miss. 1973); *Commonwealth v. Whitacre*, 878 A.2d 96, 101 (Pa. Super. Ct. 2005) (“no abuse of discretion in the trial court’s decision to permit admission of the evidence regarding comparison of the two shell casings with the shotgun owned by Appellant”).

testify that a bullet *could have been fired* from a particular firearm;²⁸⁰ that is, the class characteristics of the bullet and the firearm are consistent.²⁸¹

The early post-*Daubert* challenges to the admissibility of firearms identification evidence failed.²⁸² This changed in 2005 in *United States v. Green*,²⁸³ where the court ruled that the expert could describe only the ways in which the casings were similar but not that the casings came from a specific weapon “to the exclusion of every other firearm in the world.”²⁸⁴ In *United States v. Monteiro*²⁸⁵ the expert had not made any sketches or taken photographs and thus adequate documentation was lacking: “Until the basis for the identification is described in such a way that the procedure performed by [the examiner] is reproducible and verifiable, it is inadmissible under Rule 702.”²⁸⁶

In 2007 in *United States v. Diaz*,²⁸⁷ the court found that the record did not support the conclusion that identifications could be made to the exclusion of all other firearms in the world. Thus, “the examiners who testify in this case may only testify that a match has been made to a ‘reasonable degree of certainty in the ballistics field.’”²⁸⁸ In 2008, *United States v. Glynn*²⁸⁹ ruled that the expert could

280. *E.g.*, *People v. Horning*, 102 P.3d 228, 236 (Cal. 2004) (expert “opined that both bullets and the casing could have been fired from the same gun . . . ; because of their condition he could not say for sure”); *Luttrell v. Commonwealth*, 952 S.W.2d 216, 218 (Ky. 1997) (expert “testified only that the bullets which killed the victim could have been fired from Luttrell’s gun”); *State v. Reynolds*, 297 S.E.2d 532, 539–40 (N.C. 1982); *Commonwealth v. Moore*, 340 A.2d 447, 451 (Pa. 1975).

281. This type of evidence has some probative value and satisfies the minimal evidentiary test for logical relevancy. *See* Fed. R. Evid. 401. As one court commented, the expert’s “testimony, which established that the bullet which killed [the victim] could have been fired from the same caliber and make of gun found in the possession of [the defendant], significantly advanced the inquiry.” *Commonwealth v. Hoss*, 283 A.2d 58, 68 (Pa. 1971).

282. *See* *United States v. Hicks*, 389 F.3d 514, 526 (5th Cir. 2004) (ruling that “the matching of spent shell casings to the weapon that fired them has been a recognized method of ballistics testing in this circuit for decades”); *United States v. Foster*, 300 F. Supp. 2d 375, 377 n.1 (D. Md. 2004) (“Ballistics evidence has been accepted in criminal cases for many years. . . . In the years since *Daubert*, numerous cases have confirmed the reliability of ballistics identification.”); *United States v. Santiago*, 199 F. Supp. 2d 101, 111 (S.D.N.Y. 2002) (“The Court has not found a single case in this Circuit that would suggest that the entire field of ballistics identification is unreliable.”); *State v. Anderson*, 624 S.E.2d 393, 397–98 (N.C. Ct. App. 2006) (no abuse of discretion in admitting bullet identification evidence); *Whitacre*, 878 A.2d at 101 (“no abuse of discretion in the trial court’s decision to permit admission of the evidence regarding comparison of the two shell casings with the shotgun owned by Appellant”).

283. 405 F. Supp. 2d 104 (D. Mass. 2005).

284. *Id.* at 107. The court had followed the same approach in a handwriting case. *See* *United States v. Hines*, 55 F. Supp. 2d 62, 67 (D. Mass. 1999) (expert testimony concerning the general similarities and differences between a defendant’s handwriting exemplar and a stick-up note was admissible but not the specific conclusion that the defendant was the author).

285. 407 F. Supp. 2d 351 (D. Mass. 2006).

286. *Id.* at 374.

287. No. CR 05-00167 WHA, 2007 WL 485967 (N.D. Cal. Feb. 12, 2007).

288. *Id.* at *1.

289. 578 F. Supp. 2d 567 (S.D.N.Y. 2008).

not use the term “reasonable scientific certainty” in testifying. Rather, the expert would be permitted to testify only that it was “more likely than not” that recovered bullets and cartridge cases came from a particular weapon.

Yet other courts continued to uphold admission.²⁹⁰ By way of example, in *United States v. Williams*,²⁹¹ the Second Circuit upheld the admissibility of firearms identification evidence—bullets and cartridge casings. The opinion, however, contained some cautionary language: “We do not wish this opinion to be taken as saying that any proffered ballistic expert should be routinely admitted.”²⁹² Several cases limited testimony after the 2009 NAS Report was published.²⁹³ In the past, courts often have admitted toolmark identification evidence,²⁹⁴ includ-

290. See *United States v. Natson*, 469 F. Supp. 2d 1253, 1261 (M.D. Ga. 2007) (“According to his testimony, these toolmarks were sufficiently similar to allow him to identify Defendant’s gun as the gun that fired the cartridge found at the crime scene. He opined that he held this opinion to a 100% degree of certainty. . . . The Court also finds [the examiner’s] opinions reliable and based upon a scientifically valid methodology. Evidence was presented at the hearing that the toolmark testing methodology he employed has been tested, has been subjected to peer review, has an ascertainable error rate, and is generally accepted in the scientific community.”); *Commonwealth v. Meeks*, Nos. 2002-10961, 2003-10575, 2006 WL 2819423, at * 50 (Mass. Super. Ct. Sept. 28, 2006) (“The theory and process of firearms identification are generally accepted and reliable, and the process has been reliably applied in these cases. Accordingly, the firearms identification evidence, including opinions as to matches, may be presented to the juries for their consideration, but only if that evidence includes a detailed statement of the reasons for those opinions together with appropriate documentation.”).

291. 506 F.3d 151, 161–62 (2d Cir. 2007) (“*Daubert* did make plain that Rule 702 embodies a more liberal standard of admissibility for expert opinions than did *Frye*. . . . But this shift to a more permissive approach to expert testimony did not abrogate the district court’s gatekeeping function. Nor did it ‘grandfather’ or protect from *Daubert* scrutiny evidence that had previously been admitted under *Frye*.”) (citations omitted).

292. *Id.* at 161.

293. See *United States v. Willock*, 696 F. Supp. 2d 536, 546, 549 (D. Md. 2010) (holding, based on a comprehensive magistrate’s report, that “Sgt. Ensor shall not opine that it is a ‘practical impossibility’ for a firearm to have fired the cartridges other than the common ‘unknown firearm’ to which Sgt. Ensor attributes the cartridges.” Thus, “Sgt. Ensor shall state his opinions and conclusions without any characterization as to the degree of certainty with which he holds them.”); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1180 (D.N.M. 2009) (“[B]ecause of the limitations on the reliability of firearms identification evidence discussed above, Mr. Nichols will not be permitted to testify that his methodology allows him to reach this conclusion as a matter of scientific certainty. Mr. Nichols also will not be allowed to testify that he can conclude that there is a match to the exclusion, either practical or absolute, of all other guns. He may only testify that, in his opinion, the bullet came from the suspect rifle to within a reasonable degree of certainty in the firearms examination field.”).

294. In 1975, the Ninth Circuit noted that toolmark identification “rests upon a scientific basis and is a reliable and generally accepted procedure.” *United States v. Bowers*, 534 F.2d 186, 193 (9th Cir. 1976).

ing screwdrivers,²⁹⁵ crowbars,²⁹⁶ punches,²⁹⁷ knives,²⁹⁸ as well as other objects.²⁹⁹ An expert's opinion is admissible even if the expert cannot testify to a positive identification.³⁰⁰

IX. Bite Mark Evidence

Bite mark analysis has been used for more than 50 years to establish a connection between a defendant and a crime.³⁰¹ The specialty developed within the field of forensic dentistry as an adjunct of dental identification, rather than originating in

295. *E.g.*, *State v. Dillon*, 161 N.W.2d 738, 741 (Iowa 1968) (screwdriver and nail bar fit marks on door frame); *State v. Wessling*, 150 N.W.2d 301 (Iowa 1967) (screwdriver); *State v. Hazelwood*, 498 P.2d 607, 612 (Kan. 1972) (screwdriver and imprint on window molding); *State v. Wade*, 465 S.W.2d 498, 499–500 (Mo. 1971) (screwdriver and pry marks on door jamb); *State v. Brown*, 291 S.W.2d 615, 618–19 (Mo. 1956) (crowbar and screwdriver marks on window sash and door); *State v. Eickmeier*, 191 N.W.2d 815, 816 (Neb. 1971) (screwdriver and marks on door).

296. *E.g.*, *Brown*, 291 S.W.2d at 618–19 (Mo. 1956) (crowbar and screwdriver marks on window sash and door); *State v. Raines*, 224 S.E.2d 232, 234 (N.C. Ct. App. 1976).

297. *E.g.*, *State v. Montgomery*, 261 P.2d 1009, 1011–12 (Kan. 1953) (punch marks on safe).

298. *E.g.*, *State v. Baldwin*, 12 P. 318, 324–25 (Kan. 1886) (experienced carpenters could testify that wood panel could have been cut by accused's knife); *Graves v. State*, 563 P.2d 646, 650 (Okla. Crim. App. 1977) (blade and knife handle matched); *State v. Clark*, 287 P. 18, 20 (Wash. 1930) (knife and cuts on tree branches); *State v. Bernson*, 700 P.2d 758, 764 (Wash. Ct. App. 1985) (knife tip comparison).

299. *E.g.*, *United States v. Taylor*, 334 F. Supp. 1050, 1056–57 (E.D. Pa. 1971) (impressions on stolen vehicle and impressions made by dies found in defendant's possession), *aff'd*, 469 F.2d 284 (3d Cir. 1972); *State v. McClelland*, 162 N.W.2d 457, 462 (Iowa 1968) (pry bar and marks on "jimmied" door); *Adcock v. State*, 444 P.2d 242, 243–44 (Okla. Crim. App. 1968) (tool matched pry marks on door molding); *State v. Olsen*, 317 P.2d 938, 940 (Or. 1957) (hammer marks on the spindle of a safe).

300. For example, in *United States v. Murphy*, 996 F.2d 94 (5th Cir. 1993), an FBI expert gave limited testimony "that the tools such as the screwdriver associated with Murphy 'could' have made the marks on the ignitions but that he could not positively attribute the marks to the tools identified with Murphy." *Id.* at 99; *see also State v. Genrich*, 928 P.2d 799, 802 (Colo. App. 1996) (upholding expert testimony that three different sets of pliers recovered from the accused's house were used to cut wire and fasten a cap found in the debris from pipe bombs: "The expert's premise, that no two tools make exactly the same mark, is not challenged by any evidence in this record. Hence, the lack of a database and points of comparison does not render the opinion inadmissible.").

Although most courts have been receptive to toolmark evidence, a notable exception was *Ramirez v. State*, 810 So. 2d 836, 849–51 (Fla. 2001). In *Ramirez*, the Florida Supreme Court rejected the testimony of five experts who claimed general acceptance for a process of matching a knife with a cartilage wound in a murder victim—a type of "toolmark" comparison. Although the court applied *Frye*, it emphasized the lack of testing, the paucity of "meaningful peer review," the absence of a quantified error rate, and the lack of developed objective standards. In *Sexton v. State*, 93 S.W.3d 96 (Tex. Crim. App. 2002), an expert testified that cartridge cases from *unfired* bullets found in the appellant's apartment had distinct marks that matched fired cartridge cases found at the scene of the offense. The court ruled the testimony inadmissible: "This record qualifies Crumley as a firearms identification expert, but does not support his capacity to identify cartridge cases on the basis of magazine marks only." *Id.* at 101.

301. *See E.H. Dinkel, The Use of Bite Mark Evidence as an Investigative Aid*, 19 J. Forensic Sci. 535 (1973).

crime laboratories. Courts have admitted bite mark comparison evidence in homicide, rape, and child abuse cases. In virtually all the cases, the evidence was first offered by the prosecution. The typical bite mark case has involved the identification of the defendant by matching his dentition with a mark left on the victim. In several cases, however, the victim's teeth have been compared with marks on the defendant's body. One bite mark case involved dentures³⁰² and another braces.³⁰³ A few cases have entailed bite impressions on foodstuff found at a crime scene: apple,³⁰⁴ piece of cheese,³⁰⁵ and sandwich.³⁰⁶ Still other cases involved dog bites.³⁰⁷

Bite marks occur primarily in sex-related crimes, child abuse cases, and offenses involving physical altercations, such as homicide. A survey of 101 cases reported these findings: "More than one bitemark was present in 48% of all the bite cases studied. Bitemarks were found on adults in 81.3% of the cases and on children under 18 years-of-age in 16.7% of cases. Bitemarks were associated with the following types of crimes: murder, including attempted murder (53.9%), rape (20.8%), sexual assault (9.7%), child abuse (9.7%), burglary (3.3%), and kidnapping (12.6%)."³⁰⁸

A. The Technique

Bite mark identification is an offshoot of the dental identification of deceased persons, which is often used in mass disasters. Dental identification is based on the assumption that every person's dentition is unique. The human adult dentition consists of 32 teeth, each with 5 anatomic surfaces. Thus, there are 160 dental surfaces that can contain identifying characteristics. Restorations, with varying shapes, sizes, and restorative materials, may offer numerous additional points of individuality. Moreover, the number of teeth, prostheses, decay, malposition,

302. See *Rogers v. State*, 344 S.E.2d 644, 647 (Ga. 1986) ("Bite marks on one of Rogers' arms were consistent with the dentures worn by the elderly victim.").

303. See *People v. Shaw*, 664 N.E.2d 97, 101, 103 (Ill. App. Ct. 1996) (In a murder and aggravated sexual assault prosecution, the forensic odontologist opined that the mark on the defendant was caused by the orthodontic braces on the victim's teeth; "Dr. Kenney admitted that he was not a certified toolmark examiner"; no abuse of discretion to admit evidence).

304. See *State v. Ortiz*, 502 A.2d 400, 401 (Conn. 1985).

305. See *Doyle v. State*, 263 S.W.2d 779, 779 (Tex. Crim. App. 1954); *Seivewright v. State*, 7 P.3d 24, 26 (Wyo. 2000) ("On the basis of his comparison of the impressions from the cheese with Seivewright's dentition, Dr. Huber concluded that Seivewright was the person who bit the cheese.").

306. See *Banks v. State*, 725 So. 2d 711, 714-16 (Miss. 1997) (finding a due process violation when prosecution expert threw away sandwich after finding the accused's teeth consistent with the sandwich bite).

307. See *Davasher v. State*, 823 S.W.2d 863, 870 (Ark. 1992) (expert testified that victim's dog could be eliminated as the source of mark found on defendant); *State v. Powell*, 446 S.E.2d 26, 27-28 (N.C. 1994) ("A forensic odontologist testified that dental impressions taken from Bruno and Woody [accused's dogs] were compatible with some of the lacerations in the wounds pictured in scale photographs of Prevet's body.").

308. Iain A. Pretty & David J. Sweet, *Anatomical Location of Bitemarks and Associated Findings in 101 Cases from the United States*, 45 J. Forensic Sci. 812, 812 (2000).

malrotation, peculiar shapes, root canal therapy, bone patterns, bite relationship, and oral pathology may also provide identifying characteristics.³⁰⁹ The courts have accepted dental identification as a means of establishing the identity of a homicide victim,³¹⁰ with some cases dating back to the nineteenth century.³¹¹ According to one court, “it cannot be seriously disputed that a dental structure may constitute a means of identifying a deceased person . . . where there is some dental record of that person with which the structure may be compared.”³¹²

1. Theory of uniqueness

Identification of a suspect by matching his or her dentition with a bite mark found on the victim of a crime rests on the theory that each person’s dentition is unique. However, there are significant differences between the use of forensic dental techniques to identify a decedent and the use of bite mark analysis to identify a perpetrator.³¹³ In 1969, when bite mark comparisons were first studied, one authority raised the following problems:

[Bite]marks can never be taken to reproduce accurately the dental features of the originator. This is due partially to the fact that bite marks generally include only a limited number of teeth. Furthermore, the material (whether food stuff or human skin) in which the mark has been left is usually found to be a very unsatisfactory impression material with shrinkage and distortion characteristics that are unknown. Finally, these marks represent only the remaining and fixed picture of an action, the mechanism of which may vary from case to case. For instance, there is as yet no precise knowledge of the possible differences between biting off a morsel of food and using one’s teeth for purposes of attack or defense.³¹⁴

309. The identification is made by comparing the decedent’s teeth with antemortem dental records, such as charts and, more importantly, radiographs.

310. *E.g.*, *Wooley v. People*, 367 P.2d 903, 905 (Colo. 1961) (dentist compared his patient’s record with dentition of a corpse); *Martin v. State*, 636 N.E.2d 1268, 1272 (Ind. Ct. App. 1994) (dentist qualified to compare X rays of one of his patients with skeletal remains of murder victim and make a positive identification); *Fields v. State*, 322 P.2d 431, 446 (Okla. Crim. App. 1958) (murder case in which victim was burned beyond recognition).

311. *See Commonwealth v. Webster*, 59 Mass. (5 Cush.) 295, 299–300 (1850) (remains of the incinerated victim, including charred teeth and parts of a denture, were identified by the victim’s dentist); *Lindsay v. People*, 63 N.Y. 143, 145–46 (1875).

312. *People v. Mattox*, 237 N.E.2d 845, 846 (Ill. App. Ct. 1968).

313. *See* Iain A. Pretty & David J. Sweet, *The Scientific Basis for Human Bitemark Analyses—A Critical Review*, 41 *Sci. & Just.* 85, 88 (2001) (“A distinction must be drawn from the ability of a forensic dentist to identify an individual from their dentition by using radiographs and dental records and the science of bitemark analysis.”).

314. S. Keiser-Nielsen, *Forensic Odontology*, 1 *U. Tol. L. Rev.* 633, 636 (1969); *see also* NRC Forensic Science Report, *supra* note 3, at 174 (“[B]ite marks on the skin will change over time and can be distorted by the elasticity of the skin, the unevenness of the surface bite, and swelling and healing. These features may severely limit the validity of forensic odontology. Also, some practical difficulties, such as distortions in photographs and changes over time in the dentition of suspects, may limit the accuracy of the results.”).

Dental identifications of decedents do not pose any of these problems; the expert can often compare all 32 teeth with X rays depicting all those teeth. However, in the typical bite mark case, all 32 teeth cannot be compared; often only 4 to 8 are biting teeth that can be compared. Similarly, all five anatomic surfaces are not engaged in biting; only the edges of the front teeth come into play. In sum, bite mark identification depends not only on the uniqueness of each person's dentition but also on "whether there is a [sufficient] representation of that uniqueness in the mark found on the skin or other inanimate object."³¹⁵

2. *Methods of comparison*

Several methods of bite mark analysis have been reported. All involve three steps: (1) registration of both the bite mark and the suspect's dentition, (2) comparison of the dentition and bite mark, and (3) evaluation of the points of similarity or dissimilarity. The reproductions of the bite mark and the suspect's dentition are analyzed through a variety of methods.³¹⁶ The comparison may be either direct or indirect. A model of the suspect's teeth is used in direct comparisons; the model is compared to life-size photographs of the bite mark. Transparent overlays made from the model are used in indirect comparisons.

Although the expert's conclusions are based on objective data, the ultimate opinion regarding individuation is essentially a subjective one.³¹⁷ There is no accepted minimum number of points of identity required for a positive identification.³¹⁸ The experts who have appeared in published bite mark cases have testified to a wide range of points of similarity, from a low of eight points to a

315. Raymond D. Rawson et al., *Statistical Evidence for the Individuality of the Human Dentition*, 29 J. Forensic Sci. 252 (1984).

316. See David J. Sweet, *Human Bitemarks: Examination, Recovery, and Analysis*, in *Manual of Forensic Odontology* 162 (American Society of Forensic Odontology, 3d ed. 1997) [hereinafter *ASFO Manual*] ("The analytical protocol for bitemark comparison is made up of two broad categories. Firstly, the measurement of specific traits and features called a *metric analysis*, and secondly, the physical matching or comparison of the configuration and pattern of the injury called a *pattern association*."); see also David J. Sweet & C. Michael Bowers, *Accuracy of Bite Mark Overlays: A Comparison of Five Common Methods to Produce Exemplars from a Suspect's Dentition*, 43 J. Forensic Sci. 362, 362 (1998) ("A review of the forensic odontology literature reveals multiple techniques for overlay production. There is an absence of reliability testing or comparison of these methods to known or reference standards.").

317. See Roland F. Kouble & Geoffrey T. Craig, *A Comparison Between Direct and Indirect Methods Available for Human Bite Mark Analysis*, 49 J. Forensic Sci. 111, 111 (2004) ("It is important to remember that computer-generated overlays still retain an element of subjectivity, as the selection of the biting edge profiles is reliant on the operator placing the 'magic wand' onto the areas to be highlighted within the digitized image.").

318. See Keiser-Nielson, *supra* note 314, at 637–38; see also *Stubbs v. State*, 845 So. 2d 656, 669 (Miss. 2003) ("There is little consensus in the scientific community on the number of points which must match before any positive identification can be announced.").

high of 52 points.³¹⁹ Moreover, disagreements among experts in court appear commonplace: “Although bite mark evidence has demonstrated a high degree of acceptance, it continues to be hotly contested in ‘battles of the experts.’ Review of trial transcripts reveals that distortion and the interpretation of distortion is a factor in most cases.”³²⁰ Because of the subjectivity, some odontologists have argued that “bitemark evidence should only be used to exclude a suspect. This [argument] is supported by research which shows that the exclusion of non-biters within a population of suspects is extremely accurate; far more so than the positive identification of biters.”³²¹

3. ABFO Guidelines

In an attempt to develop an objective method, in 1984 the American Board of Forensic Odontology (ABFO) promulgated guidelines for bite mark analysis, including a uniform scoring system.³²² According to the drafting committee, “[t]he scoring system . . . has demonstrated a method of evaluation that produced a high degree of reliability among observers.”³²³ Moreover, the committee characterized “[t]he scoring guide . . . [as] the beginning of a truly scientific approach to bite mark analysis.”³²⁴ In a subsequent letter, however, the drafting committee wrote:

While the Board’s published guidelines suggest use of the scoring system, the authors’ present recommendation is that all odontologists await the results of further research before relying on precise point counts in evidentiary proceedings. . . . [T]he authors believe that further research is needed regarding the quantification of bite mark evidence before precise point counts can be relied upon in court proceedings.³²⁵

319. *E.g.*, *State v. Garrison*, 585 P.2d 563, 566 (Ariz. 1978) (10 points); *People v. Slone*, 143 Cal. Rptr. 61, 67 (Cal. Ct. App. 1978) (10 points); *People v. Milone*, 356 N.E.2d 1350, 1356 (Ill. App. Ct. 1976) (29 points); *State v. Sager*, 600 S.W.2d 541, 564 (Mo. Ct. App. 1980) (52 points); *State v. Green*, 290 S.E.2d 625, 630 (N.C. 1982) (14 points); *State v. Temple*, 273 S.E.2d 273, 279 (N.C. 1981) (8 points); *Kennedy v. State*, 640 P.2d 971, 976 (Okla. Crim. App. 1982) (40 points); *State v. Jones*, 259 S.E.2d 120, 125 (S.C. 1979) (37 points).

320. Raymond D. Rawson et al., *Analysis of Photographic Distortion in Bite Marks: A Report of the Bite Mark Guidelines Committee*, 31 J. Forensic Sci. 1261, 1261–62 (1986). The committee noted: “[P]hotographic distortion can be very difficult to understand and interpret when viewing prints of bite marks that have been photographed from unknown angles.” *Id.* at 1267.

321. Iain A. Pretty, *A Web-Based Survey of Odontologist’s Opinions Concerning Bitemark Analyses*, 48 J. Forensic Sci. 1117, 1120 (2003) [hereinafter *Web-Based Survey*].

322. ABFO, *Guidelines for Bite Mark Analysis*, 112 J. Am. Dental Ass’n 383 (1986).

323. Raymond D. Rawson et al., *Reliability of the Scoring System of the American Board of Forensic Odontology for Human Bite Marks*, 31 J. Forensic Sci. 1235, 1259 (1986).

324. *Id.*

325. Letter, *Discussion of “Reliability of the Scoring System of the American Board of Forensic Odontology for Human Bite Marks,”* 33 J. Forensic Sci. 20 (1988).

B. The Empirical Record

The 2009 NRC report concluded:

More research is needed to confirm the fundamental basis for the science of bite mark comparison. Although forensic odontologists understand the anatomy of teeth and the mechanics of biting and can retrieve sufficient information from bite marks on skin to assist in criminal investigations and provide testimony at criminal trials, the scientific basis is insufficient to conclude that bite mark comparisons can result in a conclusive match.³²⁶

Moreover, “[t]here is no science on the reproducibility of the different methods of analysis that lead to conclusions about the probability of a match.”³²⁷ Another passage provides: “Despite the inherent weaknesses involved in bite mark comparison, it is reasonable to assume that the process can sometimes reliably exclude suspects.”³²⁸

Although bitemark identifications are accepted by forensic dentists, only a few empirical studies have been conducted³²⁹ and only a small number of forensic dentists have addressed the empirical issue. In the words of one expert,

The research suggests that bitemark evidence, at least that which is used to identify biters, is a potentially valid and reliable methodology. It is generally accepted within the scientific [dental] community, although the basis of this acceptance within the peer-reviewed literature is thin. Only three studies have examined the ability of odontologists to utilise bitemarks for the identification of biters, and only two studies have been performed in what could be considered a contemporary framework of attitudes and techniques.³³⁰

326. NRC Forensic Science Report, *supra* note 3, at 175. See also *id.* at 176. (“Although the majority of forensic odontologists are satisfied that bite marks can demonstrate sufficient detail for positive identification, no scientific studies support this assessment, and no large population studies have been conducted.”),

327. *Id.* at 174.

328. *Id.* at 176.

329. See C. Michael Bowers, *Forensic Dental Evidence: An Investigator’s Handbook* 189 (2004) (“As a number of legal commentators have observed, bite mark analysis has never passed through the rigorous scientific examination that is common to most sciences. The literature does not go far in disputing that claim.”); Iain A. Pretty, *Unresolved Issues in Bitemark Analysis*, in *Bitemark Evidence* 547, 547 (Robert B.J. Dorion ed., 2005) (“As a general rule, case reports add little to the scientific knowledge base, and therefore, if these, along with noncritical reviews, are discarded, very little new empirical evidence has been developed in the past five years.”); *id.* at 561 (“[T]he final question in the recent survey asked, ‘Should an appropriately trained individual positively identify a suspect from a bitemark on skin’—70% of the respondents stated yes. However, it is the judicial system that must assess validity, reliability, and a sound scientific base for expert forensic testimony. A great deal of further research is required if odontology hopes to continue to be a generally accepted science.”).

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

Commentators have highlighted the following areas of controversy: “a) accuracy of the bitemark itself, b) uniqueness of the human dentition, and c) analytical techniques.”³³¹

One part of a 1975 study involved identification of bites made on pigskin: “Incorrect identification of the bites made on pigskin ranged from 24% incorrect identifications under ideal laboratory conditions to as high as 91% incorrect identifications when the bites were photographed 24 hours after the bites made.”³³² A 1999 ABFO Workshop, “where ABFO diplomats attempted to match four bite-marks to seven dental models, resulted in 63.5% false positives.”³³³ A 2001 study of bites on pigskin “found false positive identifications of 11.9–22.0% for various groups of forensic odontologists (15.9% false positives for ABFO diplomats), with some ABFO diplomats faring far worse.”³³⁴ Other commentators take a more favorable view of these studies.³³⁵

1. DNA exonerations

In several cases, subsequent DNA testing has demonstrated the error in a prior bite mark identification. In *State v. Krone*,³³⁶ two experienced experts concluded that the defendant had made the bite mark found on a murder victim. The defendant, however, was later exonerated through DNA testing.³³⁷ In *Otero v. Warnick*,³³⁸ a forensic dentist testified that the “plaintiff was the only person in the world who

331. Pretty & Sweet, *supra* note 313, at 87. Commentators had questioned the lack of research in the field as long ago as 1985. Two commentators wrote:

There is effectively no valid documented scientific data to support the hypothesis that bite marks are demonstrably unique. Additionally, there is no documented scientific data to support the hypothesis that a latent bite mark, like a latent fingerprint, is a true and accurate reflection of this uniqueness. To the contrary, what little scientific evidence that does exist clearly supports the conclusion that crime-related bite marks are grossly distorted, inaccurate, and therefore unreliable as a method of identification.

Allen P. Wilkinson & Ronald M. Gerugthy, *Bite Mark Evidence: Its Admissibility Is Hard to Swallow*, 12 W. St. U. L. Rev. 519, 560 (1985).

332. C. Michael Bowers, *Problem-Based Analysis of Bitemark Misidentifications: The Role of DNA*, 159S Forensic Sci. Int'l S104, S106 (2006) (citing D.K. Whittaker, *Some Laboratory Studies on the Accuracy of Bite Mark Comparison*, 25 Int'l Dent. J. 166 (1975)) [hereinafter *Problem-Based Analysis*].

333. Bowers, *Problem-Based Analysis*, *supra* note 332, at S106. *But see* Kristopher L. Arheart & Iain A. Pretty, *Results of the 4th ABFO Bitemark Workshop 1999*, 124 Forensic Sci. Int'l 104 (2001).

334. Bowers, *Problem-Based Analysis*, *supra* note 332, at S106 (citing Iain A. Pretty & David J. Sweet, *Digital Bitemark Overlays—An Analysis of Effectiveness*, 46 J. Forensic Sci. 1385, 1390 (2001) (“While the overall effectiveness of overlays has been established, the variation in individual performance of odontologists is of concern.”)).

335. See Pretty, *Reliability of Bitemark Evidence*, in *Bitemark Evidence*, *supra* note 330, at 538–42.

336. 897 P.2d 621, 622, 623 (Ariz. 1995) (“The bite marks were crucial to the State’s case because there was very little other evidence to suggest Krone’s guilt.”; “Another State dental expert, Dr. John Piakis, also said that Krone made the bite marks. . . . Dr. Rawson himself said that Krone made the bite marks. . . .”).

337. See Mark Hansen, *The Uncertain Science of Evidence*, A.B.A. J. 49 (2005) (discussing *Krone*).

338. 614 N.W.2d 177 (Mich. Ct. App. 2000).

could have inflicted the bite marks on [the murder victim's] body. On January 30, 1995, the Detroit Police Crime Laboratory released a supplemental report that concluded that plaintiff was excluded as a possible source of DNA obtained from vaginal and rectal swabs taken from [the victim's] body."³³⁹ In *Burke v. Town of Walpole*,³⁴⁰ the expert concluded that "Burke's teeth matched the bite mark on the victim's left breast to a 'reasonable degree of scientific certainty.' That same morning . . . DNA analysis showed that Burke was excluded as the source of male DNA found in the bite mark on the victim's left breast."³⁴¹ In the future, the availability of nuclear DNA testing may reduce the need to rely on bite mark identifications.³⁴²

C. Case Law Development

People v. Marx (1975)³⁴³ emerged as the leading bite mark case. After *Marx*, bite mark evidence became widely accepted.³⁴⁴ By 1992, it had been introduced or noted in 193 reported cases and accepted as admissible in 35 states.³⁴⁵ Some courts described bite mark comparison as a "science,"³⁴⁶ and several cases took judicial notice of its validity.³⁴⁷

339. *Id.* at 178.

340. 405 F.3d 66, 73 (1st Cir. 2005).

341. See also Bowers, *Problem-Based Analysis*, *supra* note 332, at S104 (citing several cases involving bitemarks and DNA exonerations: *Gates*, *Bourne*, *Morris*, *Krone*, *Otero*, *Young*, and *Brewer*); Mark Hansen, *Out of the Blue*, A.B.A. 50, 51 (1996) (DNA analysis of skin taken from fingernail scrapings of the victim conclusively excluded Bourne).

342. See Pretty, *Web-Based Survey*, *supra* note 321, at 1119 ("The use of DNA in the assessment of bitemarks has been established for some time, although previous studies have suggested that the uptake of this technique has been slow. It is encouraging to note that nearly half of the respondents in this case have employed biological evidence in a bitemark case.").

343. 126 Cal. Rptr. 350 (Cal. Ct. App. 1975). The court in *Marx* avoided applying the *Frye* test, which requires acceptance of a novel technique by the scientific community as a prerequisite to admissibility. According to the court, the *Frye* test "finds its rational basis in the degree to which the trier of fact must accept, on faith, scientific hypotheses not capable of proof or disproof in court and not even generally accepted outside the courtroom." *Id.* at 355–56.

344. Two Australian cases, however, excluded bite mark evidence. See *Lewis v. The Queen* (1987) 29 A. Crim. R. 267 (odontological evidence was improperly relied on, in that this method has not been scientifically accepted); *R v. Carroll* (1985) 19 A. Crim. R. 410 ("[T]he evidence given by the three odontologist is such that it would be unsafe or dangerous to allow a verdict based upon it to stand.").

345. Steven Weigler, *Bite Mark Evidence: Forensic Odontology and the Law*, 2 Health Matrix: J.L.-Med. 303 (1992).

346. See *People v. Marsh*, 441 N.W.2d 33, 35 (Mich. Ct. App. 1989) ("the science of bite mark analysis has been extensively reviewed in other jurisdictions"); *State v. Sager*, 600 S.W.2d 541, 569 (Mo. Ct. App. 1980) ("an exact science").

347. See *State v. Richards*, 804 P.2d 109, 112 (Ariz. Ct. App. 1990) ("[B]ite mark evidence is admissible without a preliminary determination of reliability. . . ."); *People v. Middleton*, 429 N.E.2d 100, 101 (N.Y. 1981) ("The reliability of bite mark evidence as a means of identification is sufficiently

1. Specificity of opinion

In some cases, experts testified only that a bite mark was “consistent with” the defendant’s teeth.³⁴⁸ In other cases, they went further and opined that it is “highly probable” or “very highly probable” that the defendant made the mark.³⁴⁹ In still other cases, experts made positive identifications (to the exclusion of all other persons).³⁵⁰ It is not unusual to find experts disagreeing in individual cases—often over the threshold question of whether a wound was even a bite mark.³⁵¹

established in the scientific community to make such evidence admissible in a criminal case, without separately establishing scientific reliability in each case. . . .”); *State v. Armstrong*, 369 S.E.2d 870, 877 (W. Va. 1988) (judicially noticing the reliability of bite mark evidence).

348. *E.g.*, *Rogers v. State*, 344 S.E.2d 644, 647 (Ga. 1986) (“Bite marks on one of Rogers’ arms were consistent with the dentures worn by the elderly victim.”); *People v. Williams*, 470 N.E.2d 1140, 1150 (Ill. App. Ct. 1984) (“could have”); *State v. Hodgson*, 512 N.W.2d 95, 98 (Minn. 1994) (en banc) (Board-certified forensic odontologist testified that “there were several similarities between the bite mark and the pattern of [the victim’s] teeth, as revealed by known molds of his mouth.”); *State v. Routh*, 568 P.2d 704, 705 (Or. Ct. App. 1977) (“similarity”); *Williams v. State*, 838 S.W.2d 952, 954 (Tex. Ct. App. 1992) (“One expert, a forensic odontologist, testified that Williams’s dentition was consistent with the injury (bite mark) on the deceased.”); *State v. Warness*, 893 P.2d 665, 669 (Wash. Ct. App. 1995) (“[T]he expert testified that his opinion was not conclusive, but the evidence was consistent with the alleged victim’s assertion that she had bitten Warness. . . . Its probative value was therefore limited, but its relevance was not extinguished.”).

349. *E.g.*, *People v. Slone*, 143 Cal. Rptr. 61, 67 (Cal. Ct. App. 1978); *People v. Johnson*, 289 N.E.2d 722, 726 (Ill. App. Ct. 1972).

350. *E.g.*, *Morgan v. State*, 639 So. 2d 6, 9 (Fla. 1994) (“[T]he testimony of a dental expert at trial positively matched the bite marks on the victim with Morgan’s teeth.”); *Duboise v. State*, 520 So. 2d 260, 262 (Fla. 1988) (Expert “testified at trial that within a reasonable degree of dental certainty Duboise had bitten the victim.”); *Brewer v. State*, 725 So. 2d 106, 116 (Miss. 1998) (“Dr. West opined that Brewer’s teeth inflicted the five bite mark patterns found on the body of Christine Jackson.”); *State v. Schaefer*, 855 S.W.2d 504, 506 (Mo. Ct. App. 1993) (“[A] forensic dentist testified that the bite marks on Schaefer’s shoulder matched victim’s dental impression, and concluded that victim caused the marks.”); *State v. Lyons*, 924 P.2d 802, 804 (Or. 1996) (forensic odontologist “had no doubt that the wax models were made from the same person whose teeth marks appeared on the victim’s body”); *State v. Cazes*, 875 S.W.2d 253, 258 (Tenn. 1994) (A forensic odontologist “concluded to a reasonable degree of dental certainty that Cazes’ teeth had made the bite marks on the victim’s body at or about the time of her death.”).

351. *E.g.*, *Ege v. Yukins*, 380 F. Supp. 2d 852, 878 (E.D. Mich. 2005) (“[T]he defense attempted to rebut Dr. Warnick’s testimony with the testimony of other experts who opined that the mark on the victim’s cheek was the result of *livor mortis* and was not a bite mark at all.”); *Czapleski v. Woodward*, No. C-90-0847 MHP, 1991 U.S. Dist. LEXIS 12567, at *3-4 (N.D. Cal. Aug. 30, 1991) (dentist’s initial report concluded that “bite” marks found on child were consistent with dental impressions of mother; several experts later established that the marks on child’s body were postmortem abrasion marks and not bite marks); *Kinney v. State*, 868 S.W.2d 463, 464-65 (Ark. 1994) (disagreement that marks were human bite marks); *People v. Noguera*, 842 P.2d 1160, 1165 n.1 (Cal. 1992) (“At trial, extensive testimony by forensic odontologists [sic] was presented by both sides, pro and con, as to whether the wounds were human bite marks and, if so, when they were inflicted.”); *State v. Duncan*, 802 So. 2d 533, 553 (La. 2001) (“Both defense experts testified that these marks on the victim’s body were not bite marks.”); *Stubbs v. State*, 845 So. 2d 656, 668 (Miss. 2003) (“Dr. Galvez denied the impressions found on Williams were the results of bite marks.”).

2. Post-Daubert cases

Although some commentators questioned the underlying basis for the technique after *Daubert*,³⁵² courts have continued to admit the evidence.³⁵³

X. Microscopic Hair Evidence

The first reported use of forensic hair analysis occurred more than 150 years ago in 1861 in Germany.³⁵⁴ The first published American opinion was an 1882 Wisconsin decision, *Knoll v. State*.³⁵⁵ Based on a microscopic comparison, the expert testified that the hair samples shared a common source. Hair and the closely related fiber analysis played a prominent role in two of the most famous twentieth-century American prosecutions: Ted Bundy in Florida and Wayne Williams, the alleged Atlanta child killer.³⁵⁶ Although hair comparison evidence has been judicially accepted for decades, it is another forensic identification discipline that is being reappraised today.

A. The Technique

Generally, after assessing whether a sample is a hair and not a fiber, an analyst may be able to determine: (1) whether the hair is of human or animal origin, (2) the part of the body that the hair came from, (3) whether the hair has been dyed, (4) whether the hair was pulled or fell out as a result of natural causes or disease,³⁵⁷ and (5) whether the hair was cut or crushed.³⁵⁸

352. See Pretty & Sweet, *supra* note 313, at 86 (“Despite the continued acceptance of bite-mark evidence in European, Oceanic and North American Courts the fundamental scientific basis for bite-mark analysis has never been established.”).

353. See *State v. Timmenedeas*, 737 A.2d 55, 114 (N.J. 1999) (“Judicial opinion from other jurisdictions establish that bite-mark analysis has gained general acceptance and therefore is reliable. Over thirty states considering such evidence have found it admissible and no state has rejected bite-mark evidence as unreliable.”) (citations omitted); *Stubbs*, 845 So. 2d at 670; *Howard v. State*, 853 So. 2d 781, 795–96 (Miss. 2003); *Seivewright v. State*, 7 P.3d 24, 30 (Wyo. 2000) (“Given the wide acceptance of bite mark identification testimony and Seivewright’s failure to present evidence challenging the methodology, we find no abuse of discretion in the district court’s refusal to hold an evidentiary hearing to analyze Dr. Huber’s testimony.”).

354. E. James Crocker, *Trace Evidence*, in *Forensic Evidence in Canada* 259, 265 (1991) (the analyst was Rudolf Virchow, a Berliner).

355. 12 N.W. 369 (Wis. 1882).

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

357. See *Delaware v. Fensterer*, 474 U.S. 15, 16–17 (1985) (FBI analyst testified hair found at a murder scene had been forcibly removed.).

358. See 2 Giannelli & Imwinkelried, *supra* note 177, § 24–2.

The most common subject for hair testimony involves an attempt to individuate the hair sample, at least to some degree. If the unknown is head hair, the expert might gather approximately 50 hair strands from five different areas of the scalp (the top, front, back, and both sides) from the known source.³⁵⁹ Before the microscopic analysis, the expert examines the hair macroscopically to identify obvious features visible to the naked eye such as the color of the hair and its form, that is, whether it is straight, wavy, or curved.³⁶⁰ The expert next mounts the unknown hair and the known samples on microscope slides for a more detailed examination of characteristics such as scale patterns, size, color, pigment distribution, maximum diameter, shaft length, and scale count. Some of these comparative judgments are subjective in nature: “Human hair characteristics (e.g., scale patterns, pigmentation, size) vary within a single individual. . . . Although the examination procedure involves objective methods of analysis, the subjective weights associated with the characteristics rest with the examiner.”³⁶¹

Often the examiner determines only whether the hair samples from the crime scene and the accused are “microscopically indistinguishable.” Although this finding is consistent with the hypothesis that the samples had the same source, its probative value would, of course, vary if only a hundred people had microscopically indistinguishable hair as opposed to several million. As discussed below, experts have often gone beyond this “consistent with” testimony.

B. *The Empirical Record*

The 2009 NRC report contained an assessment of hair analysis. The report began the assessment by observing that there are neither “scientifically accepted [population] frequency” statistics for various hair characteristics nor “uniform standards on the number of features on which hairs must agree before an examiner may declare a ‘match.’”³⁶² The report concluded,

[T]estimony linking microscopic hair analysis with particular defendants is highly unreliable. In cases where there seems to be a morphological match (based on microscopic examination), it must be confirmed using mtDNA analysis; microscopic studies are of limited probative value. The committee found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA. Microscopy and mtDNA analysis can be used in tandem and add to one another’s value for classifying a common source, but no studies have been performed specifically to quantify the reliability of their joint use.³⁶³

359. NRC Forensic Science Report, *supra* note 3, at 157.

360. *Id.*

361. Miller, *supra* note 67, at 157–58.

362. NRC Forensic Science Report, *supra* note 3, at 160.

363. *Id.* at 8.

There is a general consensus that hair examination can yield reliable information about class characteristics of hair strands.³⁶⁴ Indeed, experts can identify major as well as secondary characteristics. Major characteristics include such features as color, shaft form, and hair diameter.³⁶⁵ Secondary characteristics are such features as pigment size and shaft diameter.³⁶⁶ These characteristics can help narrow the class of possible sources for the unknown hair sample.

There have been several major efforts to provide an empirical basis for individuation opinions in hair analysis. In the 1940s, Gamble and Kirk investigated whether hair samples from different persons could be distinguished on the basis of scale counts.³⁶⁷ However, they used a small database of only thirty-nine hair samples, and a subsequent attempt to replicate the original experiment yielded contradictory results.³⁶⁸

In the 1960s, neutron activation analysis was used in an effort to individuate hair samples. The research focused on determining the occurrence of various trace element concentrations in human hair.³⁶⁹ Again, subsequent research tended to show that there are significant hair-to-hair variations in trace element concentration among the hairs of a single person.³⁷⁰

In the 1970s, two Canadian researchers, Gaudette and Keeping, attempted to develop a “ballpark” estimate of the probability of a false match in hair analysis. They published articles describing three studies: (1) a 1974 study involving scalp hair,³⁷¹ (2) a

364. *Id.* at 157.

365. *Id.* at 5–23.

366. *Id.*

367. Their initial research indicated that: (1) the scale count of even a single hair strand is nearly always representative of all scalp hairs; and (2) while the average or mean scale count is constant for the individual, the count differs significantly from person to person. Lucy L. Gamble & Paul L. Kirk, *Human Hair Studies II. Scale Counts*, 31 *J. Crim. L. & Criminology* 627, 629 (1941); Paul L. Kirk & Lucy L. Gamble, *Further Investigation of the Scale Count of Human Hair*, 33 *J. Crim. L. & Criminology* 276, 280 (1942).

368. Joseph Beeman, *The Scale Count of Human Hair*, 32 *J. Crim. L. & Criminology* 572, 574 (1942).

369. Rita Cornelis, *Is It Possible to Identify Individuals by Neutron Activation Analysis of Hair?* 12 *Med. Sci. & L.* 188 (1972); Lima et al., *Activation Analysis Applied to Forensic Investigation: Some Observations on the Problem of Human Hair Individualization*, 1 *Radio Chem. Methods of Analysis* 119 (Int'l Atomic Energy Agency 1965); A.K. Perkins, *Individualization of Human Head Hair*, in *Proceedings of the First Int'l Conf. on Forensic Activation Analysis* 221 (V. Guin ed., 1967).

370. Rita Cornelis, *Truth Has Many Facets: The Neutron Activation Analysis Story*, 20 *J. Forensic Sci.* 93, 95 (1980) (“I am convinced that irrefutable hair identification from its trace element composition still belongs to the realm of wishful thinking. . . . The state of the art can be said to be that nearly all interest for trace elements present in hair, as a practical identification tool, has faded.”); Dennis S. Karjala, *Evidentiary Uses of Neutron Activation Analysis*, 59 *Cal. L. Rev.* 977, 1039 (1971).

371. B.D. Gaudette & E.S. Keeping, *An Attempt at Determining Probabilities in Human Scalp Hair Comparison*, 19 *J. Forensic Sci.* 599 (1974).

1976 study using pubic hair,³⁷² and (3) a 1978 followup.³⁷³ In the two primary studies (1974 and 1976), hair samples were analyzed to determine whether hairs from different persons were microscopically indistinguishable. The analysts used 23 different characteristics such as color, pigment distribution, maximum diameter, shaft length, and scale count.³⁷⁴ Based on those data, they estimated the probability of a false match in scalp hair to be 1 in 4500 and the probability of a false match in pubic hair to be 1 in 800.

In the view of one commentator, Gaudette and Keeping's probability estimates "are easily challenged."³⁷⁵ One limitation was the relatively small database in the study.³⁷⁶ Moreover, the studies involved samples from different individuals and sought the probability that the samples from different persons would nonetheless appear microscopically indistinguishable. In a criminal trial, the question is quite different: Assuming the samples appear microscopically indistinguishable, what is the probability that they came from the same person?³⁷⁷

Early in the twenty-first century, the Verma research team revisited the individualization issue and attempted to develop an objective, automated method for identifying matches.³⁷⁸ The authors claimed that their "system accurately judged whether two populations of hairs came from the same person or from different persons 83% of the time." However, a close inspection of the authors' tabular data indicates that (1) relying on this method, researchers characterized "9 of 73 different pairs as 'same' for a false positive rate of $9/73 = 12\%$ "; and (2) the

372. B.D. Gaudette, *Probabilities and Human Pubic Hair Comparisons*, 21 J. Forensic Sci. 514, 514 (1976).

373. B.D. Gaudette, *Some Further Thoughts on Probabilities in Human Hair Comparisons*, 23 J. Forensic Sci. 758 (1978); see also Ray A. Wickenhaiser & David G. Hepworth, *Further Evaluation of Probabilities in Human Scalp Hair Comparisons*, 35 J. Forensic Sci. 1323 (1990).

374. They prescribed that with respect to each characteristic, the analysts had to classify into one of a number of specified subcategories. For example, the length characteristic was subdivided into five groups, depending on the strand's length in inches. They computed both the total number of comparisons made by the analysts and recorded the number of instances in which the analysts reported finding samples indistinguishable under the specified criteria.

375. D. Kaye, *Science in Evidence* 28 (1997); see also NRC Forensic Science Report, *supra* note 3, at 158 ([T]he "assignment of probabilities [by Gaudette and Keeping] has since been shown to be unreliable."); P.D. Barnett & R.R. Ogle, *Probabilities and Human Hair Comparisons*, 27 J. Forensic Sci. 272, 273–74 (1982); Dalva Moellenberg, *Splitting Hairs in Criminal Trials: Admissibility of Hair Comparison Probability Estimates*, 1984 Ariz. St. L.J. 521. See generally Nicholas Petrarco et al., *The Morphology and Evidential Significance of Human Hair Roots*, 33 J. Forensic Sci. 68, 68 (1988) ("Although many instrumental techniques to the individualization of human hair have been tried in recent years, these have not proved to be useful or reliable.").

376. For example, the pubic hair study involved a total of 60 individuals. In addition, the experiments involved primarily Caucasians. While the scalp hair study included 92 Caucasians, there were only 6 Asians and 2 African Americans in the study.

377. A Tawshunsky, *Admissibility of Mathematical Evidence in Criminal Trials*, 21 Am. Crim. L. Rev. 55, 57–66 (1983).

378. M.S. Verma et al., *Hair-MAP: A Prototype Automated System for Forensic Hair Comparison and Analysis*, 129 Forensic Sci. Int'l 168 (2002).

researchers characterized “4 sets of hairs from the same person as ‘different’ for a false negative rate of $4/9 = 44\%$.”³⁷⁹

The above studies do not provide the only data relevant to the validity of hair analysis. There are also comparative studies of microscopic analysis and mtDNA, proficiency tests, and DNA exoneration cases involving microscopic analysis.

1. Mitochondrial DNA³⁸⁰

An FBI study compared microscopic (“consistent with” testimony) and mtDNA analysis of hair: “Of the 80 hairs that were microscopically associated, nine comparisons were excluded by mtDNA analysis.”³⁸¹

2. Proficiency testing

Early proficiency tests indicated a high rate of laboratory error in microscopic comparisons of hair samples. In the 1970s the LEAA conducted its Laboratory Proficiency Testing Program.³⁸² The crime laboratories’ performance on hair analysis was the weakest. Fifty-four percent misanalyzed hair sample C and 67% submitted unacceptable responses on hair sample D.³⁸³ Followup studies between 1980 and 1991 yielded similar results.³⁸⁴ Summarizing the results of this series of tests, two commentators concluded: “Animal and human (body area) hair identifications are clearly the most troublesome of all categories tested.”³⁸⁵

In another series of hair tests, the examiners were asked to “include” or “exclude” in comparing known and unknown samples: “Laboratories reported inclusions and exclusions which agreed with the manufacturer in approximately 74% of their comparisons. About 18% of the responses were inconclusive, and 8% in disagreement with the manufacturers’ information.”³⁸⁶

379. NRC Forensic Science Report, *supra* note 3, at 159.

380. For a detailed discussion of mitochondrial DNA, see David H. Kaye & George Sensabaugh, Reference Guide on DNA Identification Evidence, Section V.A, in this manual

381. Max M. Houck & Bruce Budowle, *Correlation of Microscopic and Mitochondrial DNA Hair Comparisons*, 47 J. Forensic Sci. 964, 966 (2002).

382. Laboratory Proficiency Test, *supra* note 81.

383. *Id.* at 251. By way of comparison, 20% of the laboratories failed a paint analysis (test #5); 30% failed glass analysis (test #9).

384. Peterson & Markham, *supra* note 82, at 1007 (“In sum, laboratories were no more successful in identifying the correct species of origin of animal hair . . . than they were in the earlier LEAA study.”).

385. *Id.*

386. *Id.* at 1023; *see also id.* at 1022 (“Examiners warned that they needed to employ particular caution in interpreting the hair results given the virtual impossibility of achieving complete sample homogeneity.”).

3. DNA exonerations

The publication of the Department of Justice study of the first 28 DNA exonerations spotlighted the significant role that hair analysis played in several of these miscarriages of justice.³⁸⁷ For example, in the trial of Edward Honeker, an expert testified that the crime scene hair sample “was unlikely to match anyone” else³⁸⁸—a clear overstatement. Moreover, an exoneration in Canada triggered a judicial inquiry, which recommended that “[t]rial judges should undertake a more critical analysis of the admissibility of hair comparison evidence as circumstantial evidence of guilt.”³⁸⁹ One study of 200 DNA exoneration cases reported that hair testimony had been presented at 43 of the original trials.³⁹⁰ A subsequent examination of 137 trial transcripts in exoneration cases concluded: “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.”³⁹¹ The other cases contained flawed probability testimony.

C. Case Law Development

Prior to *Daubert*, an overwhelming majority of courts accepted expert testimony that hair samples are microscopically indistinguishable.³⁹² Experts often conceded that microscopic analysis did not permit a positive identification of the source.³⁹³

387. Edward Connors et al., *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial* (1996). See *id.* at 73 (discussing David Vasquez case); *id.* at 64–65 (discussing Steven Linscott case).

388. Barry Scheck et al., *Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted* 146 (2000).

389. Hon. Fred Kaufman, *The Commission on Proceedings Involving Guy Paul Morin* (Ontario Ministry of the Attorney General 1998) (Recommendation 2). Morin was erroneously convicted based, in part, on hair evidence.

390. Brandon L. Garrett, *Judging Innocence*, 108 *Colum. L. Rev.* 55, 81 (2008).

391. Garrett & Neufeld, *supra* note 33, at 47.

392. See, e.g., *United States v. Hickey*, 596 F.2d 1082, 1089 (1st Cir. 1979); *United States v. Brady*, 595 F.2d 359, 362–63 (6th Cir. 1979); *United States v. Cyphers*, 553 F.2d 1064, 1071–73 (7th Cir. 1977); *Jent v. State*, 408 So. 2d 1024, 1028–29 (Fla. 1981); *Commonwealth v. Tarver*, 345 N.E.2d 671, 676–77 (Mass. 1975); *State v. White*, 621 S.W.2d 287, 292–93 (Mo. 1981); *State v. Smith*, 637 S.W.2d 232, 236 (Mo. Ct. App. 1982); *People v. Allweiss*, 396 N.E.2d 735 (N.Y. 1979); *State v. Green*, 290 S.E.2d 625, 629–30 (N.C. 1982); *State v. Watley*, 788 P.2d 375, 381 (N.M. 1989).

393. *Moore v. Gibson*, 195 F.3d 1152, 1167 (10th Cir. 1999); *Butler v. State*, 108 S.W.3d 18, 21 (Mo. Ct. App. 2003); see also *Thompson v. State*, 539 A.2d 1052, 1057 (Del. 1988) (“it is now universally recognized that although fingerprint comparisons can result in the positive identification of an individual, hair comparisons are not this precise”). But see *People v. Kosters*, 467 N.W.2d 311, 313 (Mich. 1991) (Cavanaugh, C.J., dissenting) (the “minuscule probative value” of such opinions is “clearly . . . outweighed by the unfair prejudicial effect”); *State v. Wheeler*, 1981 WL 139588, at *4 (Wis. Ct. App. Feb. 8, 1981) (in an unpublished opinion, the appellate court held that the trial judge did not err in finding that the expert’s opinion that the accused “could have been the source” of the hair lacked probative value, because it “only include[d] defendant in a broad class of possible assailants”).

Nonetheless, the courts varied in how far they permitted the expert to go. In some cases, analysts testified only that the samples matched³⁹⁴ or were similar³⁹⁵ and thus consistent with the hypothesis that the samples had the same source.³⁹⁶ Other courts permitted experts to directly opine that the accused was the source of the crime scene sample.³⁹⁷ However, a 1990 decision held it error to admit testimony that “it would be improbable that these hairs would have originated from another individual.”³⁹⁸ In the court’s view, this testimony amounted “effectively, [to] a positive identification of defendant. . . .”³⁹⁹

On the basis of Gaudette and Keeping research, several courts admitted opinions in statistical terms (e.g., 1 in 4500 chance of a false match).⁴⁰⁰ In contrast, other courts, including a federal court of appeals, reached a contrary conclusion.⁴⁰¹

The most significant post-*Daubert* challenge to microscopic hair analysis came in *Williamson v. Reynolds*,⁴⁰² a habeas case decided in 1995. There, an expert testified that, after considering approximately 25 characteristics, he concluded that the hair samples were “consistent microscopically.” He then elaborated: “In other words, hairs are not an absolute identification, but they either came from this individual or there is—could be another individual *somewhere in the world* that would have the same characteristics to their hair.”⁴⁰³ The district court was “unsuccessful in its attempts to locate *any* indication that expert hair comparison testimony

394. *Garland v. Maggio*, 717 F.2d 199, 207 n.9 (5th Cir. 1983).

395. *United States v. Brady*, 595 F.2d 359, 362–63 (6th Cir. 1979).

396. *People v. Allen*, 115 Cal. Rptr. 839, 842 (Cal. Ct. App. 1974).

397. In the 1986 Mississippi prosecution of Randy Beville for murder, the expert testified that “there was a transfer of hair from the Defendant to the body of” the victim. Clive A. Stafford Smith & Patrick D. Goodman, *Forensic Hair Comparison Analysis: Nineteenth Century Science or Twentieth Century Snake Oil?* 27 Colum. Hum. Rts. L. Rev. 227, 273 (1996).

398. *State v. Faircloth*, 394 S.E.2d 198, 202–03 (N.C. Ct. App. 1990).

399. *Id.* at 202.

400. *United States v. Jefferson*, 17 M.J. 728, 731 (N.M.C.M.R. 1983); *People v. DiGiacomo*, 388 N.E.2d 1281, 1283 (Ill. App. Ct. 1979); *see also* *United States ex rel. DiGiacomo v. Franzen*, 680 F.2d 515, 516 (7th Cir. 1982) (During its deliberations, the jury submitted the following question to the judge: “Has it been established by sampling of hair specimens that the defendant was positively proven to have been in the automobile?”).

401. *United States v. Massey*, 594 F.2d 676, 679–80 (8th Cir. 1979) (the expert testified that he “had microscopically examined 2,000 cases and in only one or two cases was he ever unable to make identification”; the expert cited a study for the proposition that there was a 1 in 4500 chance of a random match; the expert added 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, 176 (Minn. 1978).

402. 904 F. Supp. 1529, 1554 (E.D. Okla. 1995), *rev’d on this issue sub nom. Williamson v. Ward*, 110 F.3d 1508, 1523 (10th Cir. 1997). The district court noted that the “expert did not explain which of the ‘approximately’ 25 characteristics were consistent, any standards for determining whether the samples were consistent, how many persons could be expected to share this same combination of characteristics, or how he arrived at his conclusions.” *Id.* at 1554.

403. *Id.* (emphasis added).

meets *any* of the requirements of *Daubert*.⁴⁰⁴ Finally, the prosecutor in closing argument declared, “There’s a match.”⁴⁰⁵ Even the state court had misinterpreted the evidence, writing that the “hair evidence placed [petitioner] at the decedent’s apartment.”⁴⁰⁶ Although the Tenth Circuit did not fault the district judge’s reading of the empirical record relating to hair analysis and ultimately upheld habeas relief, that court reversed the district judge on this issue. The Tenth Circuit ruled that the district had committed legal error because the due process (fundamental fairness), not the more stringent *Daubert* (reliability), standard controls evidentiary issues in habeas corpus proceedings.⁴⁰⁷ Before retrial, the defendant was exonerated by exculpatory DNA evidence.⁴⁰⁸

Post-*Daubert*, many cases have continued to admit testimony about microscopic hair analysis.⁴⁰⁹ In 1999, one state court judicially noticed the reliability of hair evidence,⁴¹⁰ implicitly finding this evidence to be not only admissible but also based on a technique of indisputable validity.⁴¹¹ In contrast, a Missouri court reasoned that, without the benefit of population frequency data, an expert overreached in opining to “a reasonable degree of certainty that the unidentified hairs were in fact from” the defendant.⁴¹² The NRC report commented that there appears to be growing judicial support for the view that “testimony linking microscopic hair analysis with particular defendants is highly unreliable.”⁴¹³

404. *Id.* at 1558. The court also observed: “Although the hair expert may have followed procedures accepted in the community of hair experts, the human hair comparison results in this case were, nonetheless, scientifically unreliable.” *Id.*

405. *Id.* at 1557.

406. *Id.* (quoting *Williamson v. State*, 812 P.2d 384, 387 (Okla. Crim. 1991)).

407. *Williamson v. Ward*, 110 F.3d 1508, 1523 (10th Cir. 1997).

408. Scheck et al., *supra* note 388, at 146 (hair evidence was shown to be “patently unreliable.”); see also John Grisham, *The Innocent Man* (2006) (examining the *Williamson* case).

409. *E.g.*, *State v. Fukusaku*, 946 P.2d 32, 44 (Haw. 1997) (“Because the scientific principles and procedures underlying hair and fiber evidence are well-established and of proven reliability, the evidence in the present case can be treated as ‘technical knowledge.’ Thus, an independent reliability determination was unnecessary.”); *McGrew v. State*, 682 N.E.2d 1289, 1292 (Ind. 1997) (concluding that hair comparison is “more a ‘matter of observation by persons with specialized knowledge’ than ‘a matter of scientific principles’”); see also NRC Forensic Science Report, *supra* note 3, at 161 n.88 (citing *State v. West*, 877 A.2d 787 (Conn. 2005), and *Bookins v. State*, 922 A.2d 389 (Del. Super. Ct. 2007)).

410. See *Johnson v. Commonwealth*, 12 S.W.3d 258, 262 (Ky. 1999).

411. See Fed. R. Evid. 201(b); *Daubert*, 509 U.S. at 593 n.11 (“[T]heories that are so firmly established as to have attained the status of scientific law, such as the laws of thermodynamics, properly are subject to judicial notice under Federal Rule [of] Evidence 201.”).

412. *Butler v. State*, 108 S.W.3d 18, 21–22 (Mo. Ct. App. 2003).

413. NRC Forensic Science Report, *supra* note 3, at 161.

XI. Recurrent Problems

The discussions of specific techniques in this chapter, as well as the 2009 NRC report, reveal several recurrent problems in the presentation of testimony about forensic expertise.

A. Clarity of Testimony

As noted earlier, the report voiced concern about the use of terms such as “match,” “consistent with,” “identical,” “similar in all respects tested,” and “cannot be excluded as the source of.” These terms can have “a profound effect on how the trier of fact in a criminal or civil matter perceives and evaluates scientific evidence.”⁴¹⁴

The comparative bullet lead cases are illustrative of this point.⁴¹⁵ The technique was used when conventional firearms identification was not possible because the recovered bullet was so deformed that the striations were destroyed. In the bullet lead cases, the phrasing of the experts’ opinions varied widely. In some, experts testified only to the limited opinion that two exhibits were “analytically indistinguishable.”⁴¹⁶ In other cases, examiners concluded that samples *could have* come from the same “source” or “batch.”⁴¹⁷ In still others, they stated that the samples *came* from the same source.⁴¹⁸ In several cases, the experts went even further and identified a particular “box” of ammunition (usually 50 loaded cartridges, sometimes 20) as the source of the bullet recovered at the crime scene. For example, experts opined that two specimens:

- Could have come from the same box.⁴¹⁹
- Could have come from the same box or a box manufactured on the same day.⁴²⁰

414. *Id.* at 21.

415. The technique compared trace chemicals found in bullets at crime scenes with ammunition found in the possession of a suspect. It was used when firearms (“ballistics”) identification could not be employed. FBI experts used various analytical techniques (first, neutron activation analysis, and then inductively coupled plasma-atomic emission spectrometry) to determine the concentrations of seven elements—arsenic, antimony, tin, copper, bismuth, silver, and cadmium—in the bullet lead alloy of both the crime-scene and suspect’s bullets. Statistical tests were then used to compare the elements in each bullet and determine whether the fragments and suspect’s bullets were “analytically indistinguishable” for each of the elemental concentration means.

416. See *Wilkerson v. State*, 776 A.2d 685, 689 (Md. Ct. Spec. App. 2001).

417. See *State v. Krummacher*, 523 P.2d 1009, 1012–13 (Or. 1974) (en banc).

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

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

420. See *State v. Grube*, 883 P.2d 1069, 1078 (Idaho 1994); *People v. Johnson*, 499 N.E.2d 1355, 1366 (Ill. 1986); *Earhart v. State*, 823 S.W.2d 607, 614 (Tex. Crim. App. 1991) (en banc) (“He

- Were consistent with their having come from the same box of ammunition.⁴²¹
- Probably came from the same box.⁴²²
- Must have come from the same box or from another box that would have been made by the same company on the same day.⁴²³

Moreover, these inconsistent statements were not supported by empirical research. According to a 2004 NRC report, the number of bullets that can be produced from an “analytically indistinguishable” melt “can range from the equivalent of as few as 12,000 to as many as 35 million 40 grain, .22 caliber long-rifle bullets.”⁴²⁴ Consequently, according to the 2004 NRC report, the “available data do not support any statement that a crime bullet came from a particular box of ammunition. [R]eferences to ‘boxes’ of ammunition in any form should be excluded as misleading under Federal Rule of Evidence 403.”⁴²⁵

B. Limitations on Testimony

Some courts have limited the scope of the testimony, permitting expert testimony about the similarities and dissimilarities between exemplars but not the specific conclusion that the defendant was the author (“common authorship” opinion).⁴²⁶ Although the courts have used this approach most frequently in questioned docu-

later modified that statement to acknowledge that analytically indistinguishable bullets which do not come from the same box most likely would have been manufactured at the same place on or about the same day; that is, in the same batch.”), *vacated*, 509 U.S. 917 (1993).

421. *See* *State v. Reynolds*, 297 S.E.2d 532, 534 (N.C. 1982).

422. *See* *Bryan v. State*, 935 P.2d 338, 360 (Okla. Crim. App. 1997).

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

424. National Research Council, *Forensic Analysis: Weighing Bullet Lead Evidence* 6 (2004), [hereinafter *NRC Bullet Lead Evidence*], available at http://www.nap.edu/catalog.php?record_id=10924.

425. *Id.*

426. *See* *United States v. Oskowitz*, 294 F. Supp. 2d 379, 384 (E.D.N.Y. 2003) (“Many other district courts have similarly permitted a handwriting expert to analyze a writing sample for the jury without permitting the expert to offer an opinion on the ultimate question of authorship.”); *United States v. Rutherford*, 104 F. Supp. 2d 1190, 1194 (D. Neb. 2000) (“[T]he Court concludes that FDE Rauscher’s testimony meets the requirements of Rule 702 to the extent that he limits his testimony to identifying and explaining the similarities and dissimilarities between the known exemplars and the questioned documents. FDE Rauscher is precluded from rendering any ultimate conclusions on authorship of the questioned documents and is similarly precluded from testifying to the degree of confidence or certainty on which his opinions are based.”); *United States v. Hines*, 55 F. Supp. 2d 62, 69 (D. Mass. 1999) (expert testimony concerning the general similarities and differences between a defendant’s handwriting exemplar and a stick-up note was admissible but not the specific conclusion that the defendant was the author).

ment cases, they have sometimes applied the same approach to other types of forensic expertise such as firearms examination as well.⁴²⁷

The NRC report criticized “exaggerated”⁴²⁸ testimony such as claims of perfect accuracy,⁴²⁹ infallibility,⁴³⁰ or a zero error rate.⁴³¹ Several courts have barred excessive expert claims for lack of empirical support. For example, in *United States v. Mitchell*,⁴³² the court commented: “Testimony at the *Daubert* hearing indicated that some latent fingerprint examiners insist that there is no error rate associated with their activities. . . . This would be out-of-place under Rule 702.”⁴³³ Similarly, in a firearms identification case, one court noted that

during the testimony at the hearing, the examiners testified to the effect that they could be 100 percent sure of a match. Because an examiner’s bottom line opinion as to an identification is largely a subjective one, there is no reliable statistical or scientific methodology which will currently permit the expert to testify that it is a ‘match’ to an absolute certainty, or to an arbitrary degree of statistical certainty.⁴³⁴

Other courts have excluded the use of terms such as “science” or “scientific,” because of the risk that jurors may bestow the aura of the infallibility of science on the testimony.⁴³⁵

In particular, some courts are troubled by the use of the expression “reasonable scientific certainty” by some forensic experts. The term “reasonable scientific certainty” is problematic. Although it is used frequently in cases, its legal meaning is ambiguous.⁴³⁶ Sometimes it is used in lieu of a confidence statement (i.e., “high degree of certainty”), in which case the expert could altogether avoid the term and directly testify how confident he or she is in the opinion.

In other cases, courts have interpreted reasonable scientific certainty to mean that the expert must testify that a sample *probably* came from the defendant and not

427. *United States v. Green*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005).

428. NRC Forensic Science Report, *supra* note 3, at 4.

429. *Id.* at 47.

430. *Id.* at 104.

431. *Id.* at 142–43.

432. 365 F.3d 215 (3d Cir. 2004).

433. *Id.* at 246.

434. *United States v. Monteiro*, 407 F. Supp. 2d 351, 372 (D. Mass. 2006).

435. *United States v. Starzeczyzel*, 880 F. Supp. 1027, 1038 (S.D.N.Y. 1995).

436. James E. Hullverson, *Reasonable Degree of Medical Certainty: A Tort et a Travers*, 31 St. Louis U. L.J. 577, 582 (1987) (“[T]here is nevertheless an undercurrent that the expert in federal court express some basis for both the confidence with which his conclusion is formed, and the probability that his conclusion is accurate.”); Edward J. Imwinkelried & Robert G. Scofield, *The Recognition of an Accused’s Constitutional Right to Introduce Expert Testimony Attacking the Weight of Prosecution Science Evidence: The Antidote for the Supreme Court’s Mistaken Assumption in California v. Trombetta*, 33 Ariz. L. Rev. 59, 69 (1991) (“Many courts continue to exclude opinions which fall short of expressing a probability or certainty. . . . These opinions have been excluded in jurisdictions which have adopted the Federal Rules of Evidence.”).

that it *possibly* came from the defendant.⁴³⁷ However, experts frequently testify that two samples “could have come from the same source.” Such testimony meets the relevancy standard of Federal Rule 401, and there is no requirement in Article VII of the Federal Rules that an expert’s opinion be expressed in terms of “probabilities.” Thus, in *United States v. Cyphers*⁴³⁸ the expert testified that hair samples found on items used in a robbery “could have come” from the defendants.⁴³⁹ The defendants argued that the testimony was inadmissible because the expert did not express his opinion in terms of reasonable scientific certainty. The court wrote: “There is no such requirement.”⁴⁴⁰

In *Burke v. Town of Walpole*,⁴⁴¹ a bite mark identification case, the court of appeals had to interpret the term as used in an arrest warrant:

[W]e must assume that the magistrate who issued the arrest warrant assigned no more than the commonly accepted meaning among lawyers and judges to the term “reasonable degree of scientific certainty”—“a standard requiring a showing that the injury was more likely than not caused by a particular stimulus, based on the general consensus of recognized [scientific] thought.” Black’s Law Dictionary 1294 (8th ed. 2004) (defining “reasonable medical probability,” or “reasonable medical certainty,” as used in tort actions). That standard, of course, is fully consistent with the probable cause standard.⁴⁴²

The case involved the guidelines adopted by ABFO that recognized several levels of certainty (“reasonable medical certainty,” “high degree of certainty,” and “virtual certainty”). The guidelines described “reasonable medical certainty” as “convey[ing] the connotation of virtual certainty or beyond reasonable doubt.”⁴⁴³ This is not the way that some courts use the term.

437. *State v. Holt*, 246 N.E.2d 365, 368 (Ohio 1969). The expert testified, based on neutron activation analysis, that two hair samples were “similar *and . . . likely* to be from the same source” (emphasis in original).

438. 553 F.2d 1064 (7th Cir. 1977).

439. *Id.* at 1072; *see also* *United States v. Davis*, 44 M.J. 13, 16 (C.A.A.F. 1996) (“Evidence was also admitted that appellant owned sneakers which ‘could have’ made these prints.”).

440. *Cyphers*, 553 F.2d at 1072; *see also* *United States v. Oaxaca*, 569 F.2d 518, 526 (9th Cir. 1978) (expert’s opinion regarding hair comparison admissible even though expert was less than certain); *United States v. Spencer*, 439 F.2d 1047, 1049 (2d Cir. 1971) (expert’s opinion regarding handwriting comparison admissible even though expert did not make a positive identification); *United States v. Longfellow*, 406 F.2d 415, 416 (4th Cir. 1969) (expert’s opinion regarding paint comparison admissible, even though expert did not make a positive identification); *State v. Boyer*, 406 So. 2d 143, 148 (La. 1981) (reasonable scientific certainty not required where expert testifies concerning the presence of gunshot residue based on neutron activation analysis).

441. 405 F.3d 66 (1st Cir. 2005).

442. *Id.* at 91.

443. *Id.* at 91 n.30 (emphasis omitted).

Moreover, the term may be problematic for a different reason—misleading the jury. One court ruled that the term “reasonable scientific certainty” could not be used because of the subjective nature of the opinion.⁴⁴⁴

C. Restriction of Final Argument

In a number of cases, in summation counsel has overstated the content of the expert testimony. In *People v. Linscott*,⁴⁴⁵ for example, “the prosecutor argued that hairs found in the victim’s apartment and on the victim’s body were in fact defendant’s hairs.”⁴⁴⁶ Reversing, the Illinois Supreme Court wrote: “With these statements, the prosecutor improperly argued that the hairs removed from the victim’s apartment were conclusively identified as coming from defendant’s head and pubic region. There simply was no testimony at trial to support these statements. In fact, [the prosecution experts] and the defense hair expert . . . testified that no such identification was possible.”⁴⁴⁷ DNA testing exculpated Linscott.⁴⁴⁸ Trial judges can police the attorneys’ descriptions of the testimony during closing argument as well as the content of expert testimony presented.

XII. Procedural Issues

The *Daubert* standard operates in a procedural setting, not a vacuum. In *Daubert*, the Supreme Court noted that “[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.”⁴⁴⁹ Adversarial testing presupposes advance notice of the content of the expert’s testimony and access to comparable expertise to evaluate that testimony. This section discusses some of the procedural mechanisms that trial judges may use to assure that jurors properly evaluate any expert testimony by forensic identification experts.

444. *United States v. Glynn*, 578 F. Supp. 2d 567, 568–75 (S.D.N.Y. 2008) (firearms identification case).

445. 566 N.E.2d 1355 (Ill. 1991).

446. *Id.* at 1358.

447. *Id.* at 1359.

448. See *Connors et al.*, *supra* note 387, at 65 (“The State’s expert on the hair examination testified that only 1 in 4,500 persons would have consistent hairs when tested for 40 different characteristics. He only tested between 8 and 12 characteristics, however, and could not remember which ones. The appellate court ruled on July 29, 1987, that his testimony, coupled with the prosecution’s use of it at closing arguments, constituted denial of a fair trial.”) (citation omitted).

449. 509 U.S. at 596 (citing *Rock v. Arkansas*, 483 U.S. 44, 61 (1987)).

A. Pretrial Discovery

Judges can monitor discovery in scientific evidence cases to ensure that disclosure is sufficiently comprehensive.⁴⁵⁰ Federal Rule 16 requires discovery of laboratory reports⁴⁵¹ and a summary of the expert's opinion.⁴⁵² The efficacy of these provisions depends on the content of the reports and the summary. The *Journal of Forensic Sciences*, the official publication of the American Academy of Forensic Sciences, published a symposium on the ethical responsibilities of forensic scientists in 1989. One symposium article described a number of unacceptable laboratory reporting practices, including (1) "preparation of reports containing minimal information in order not to give the 'other side' ammunition for cross-examination," (2) "reporting of findings without an interpretation on the assumption that if an interpretation is required it can be provided from the witness box," and (3) "[o]mitting some significant point from a report to trap an unsuspecting cross-examiner."⁴⁵³

NRC has recommended extensive discovery in DNA cases: "All data and laboratory records generated by analysis of DNA samples should be made freely available to all parties. Such access is essential for evaluating the analysis."⁴⁵⁴ The NRC report on bullet lead contained similar comments about the need for a thorough report in bullet lead cases:

The conclusions in laboratory reports should be expanded to include the limitations of compositional analysis of bullet lead evidence. In particular, a further

450. See Fed. R. Crim. P. 16 (1975) advisory committee's note ("[I]t is difficult to test expert testimony at trial without advance notice and preparation."), reprinted in 62 F.R.D. 271, 312 (1974); Paul C. Giannelli, *Criminal Discovery, Scientific Evidence, and DNA*, 44 Vand. L. Rev. 791 (1991). "Early disclosure can have the following benefits: [1] Avoiding surprise and unnecessary delay. [2] Identifying the need for defense expert services. [3] Facilitating exoneration of the innocent and encouraging plea negotiations if DNA evidence confirms guilt." National Institute of Justice, President's DNA Initiative: Principles of Forensic DNA for Officers of the Court (2005), available at <http://www.dna.gov/training/otc>.

451. Fed. R. Crim. P. 16(a)(1)(F).

452. *Id.* 16(a)(1)(G).

453. Douglas M. Lucas, *The Ethical Responsibilities of the Forensic Scientist: Exploring the Limits*, 34 J. Forensic Sci. 719, 724 (1989). Lucas was the Director of The Centre of Forensic Sciences, Ministry of the Solicitor General, Toronto, Ontario.

454. National Research Council, *DNA Technology in Forensic Science* 146 (1992) ("The prosecutor has a strong responsibility to reveal fully to defense counsel and experts retained by the defendant all material that might be necessary in evaluating the evidence."); see also *id.* at 105 ("Case records—such as notes, worksheets, autoradiographs, and population databanks—and other data or records that support examiners' conclusions are prepared, retained by the laboratory, and made available for inspection on court order after review of the reasonableness of a request."); National Research Council, *The Evaluation of Forensic DNA Evidence* 167–69 (1996) ("Certainly, there are no strictly scientific justifications for withholding information in the discovery process, and in Chapter 3 we discussed the importance of full, written documentation of all aspects of DNA laboratory operations. Such documentation would facilitate technical review of laboratory work, both within the laboratory and by outside experts. . . . Our recommendations that all aspects of DNA testing be fully documented is most valuable when this documentation is discoverable in advance of trial.").

explanatory comment should accompany the laboratory conclusions to portray the limitations of the evidence. Moreover, a section of the laboratory report translating the technical conclusions into language that a jury could understand would greatly facilitate the proper use of this evidence in the criminal justice system. Finally, measurement data (means and standard deviations) for all of the crime scene bullets and those deemed to match should be included.⁴⁵⁵

As noted earlier, the recent NRC report made similar comments:

Some reports contain only identifying and agency information, a brief description of the evidence being submitted, a brief description of the types of analysis requested, and a short statement of the results (e.g., “the greenish, brown plant material in item #1 was identified as marijuana”), and they include no mention of methods or any discussion of measurement uncertainties.⁴⁵⁶

*Melendez-Diaz v. Massachusetts*⁴⁵⁷ illustrates the problem. The laboratory report in that case “contained only the bare-bones statement that ‘[t]he substance was found to contain: Cocaine.’ At the time of trial, petitioner did not know what tests the analysts performed, whether those tests were routine, and whether interpreting their results required the exercise of judgment or the use of skills that the analysts may not have possessed.”⁴⁵⁸

1. Testifying beyond the report

Experts should generally not be allowed to testify beyond the scope of the report without issuing a supplemental report. *Troedel v. Wainwright*,⁴⁵⁹ a capital murder case, illustrates the problem. In that case, a report of a gunshot residue test based on neutron activation analysis stated the opinion that swabs “from the hands of Troedel and Hawkins contained antimony and barium in amounts typically found on the hands of a person who has discharged a firearm or has had his hands in close proximity to a discharging firearm.”⁴⁶⁰ An expert testified consistently with this report at Hawkins’ trial but embellished his testimony at Troedel’s trial by adding the more inculpatory opinion that “Troedel had fired the murder weapon.”⁴⁶¹ In contrast, at a deposition during federal habeas proceedings, the *same* expert testified that “he could not, from the results of his tests, determine or say to a scientific certainty who had fired the murder weapon” and the “amount of barium and antimony on the hands of Troedel and Hawkins were basically insignificant.”⁴⁶² The district court found the trial testimony, “at the very least,” misleading and

455. See NRC Bullet Lead Evidence, *supra* note 424, at 110–11.

456. NRC Forensic Science Report, *supra* note 3, at 21.

457. 129 S. Ct. 2527 (2009).

458. *Id.* at 2537.

459. 667 F. Supp. 1456 (S.D. Fla. 1986), *aff’d*, 828 F.2d 670 (11th Cir. 1987).

460. *Id.* at 1458.

461. *Id.*

462. *Id.* at 1459.

granted relief.⁴⁶³ The expert claimed that the prosecutor had “pushed” him to embellish his testimony, a claim the prosecutor substantiated.⁴⁶⁴

B. Defense Experts

In appropriate cases, trial judges can provide the opposition with access to expert resources. Defense experts are often important in cases involving forensic identification expertise. Counsel will frequently need expert guidance to determine whether a research study is methodologically sound and, if so, whether the data adequately support the specific opinion proffered, and the role, if any, that subjective judgment played in forming the opinion.

The NAS 1992 DNA report stressed that experts are necessary for an adequate defense in many cases: “Defense counsel must have access to adequate expert assistance, even when the admissibility of the results of analytical techniques is not in question because there is still a need to review the quality of the laboratory work and the interpretation of results.”⁴⁶⁵ According to the *President’s DNA Initiative*, “[e]ven if DNA evidence is admitted, there still may be disagreement about its interpretation—what do the DNA results mean in a particular case?”⁴⁶⁶

The need for defense experts is not limited to cases involving DNA evidence. In *Ake v. Oklahoma*,⁴⁶⁷ the Supreme Court recognized a due process right to a defense expert under certain circumstances.⁴⁶⁸ In federal trials, the Criminal Justice Act of 1964⁴⁶⁹ provides for expert assistance for indigent defendants.

463. “[T]he Court concludes that the opinion Troedel had fired the weapon was known by the prosecution not to be based on the results of the neutron activation analysis tests, or on any scientific certainty or even probability. Thus, the subject testimony was not only misleading, but also was used by the State knowing it to be misleading.” *Id.* at 1459–60.

464. *Id.* at 1459 (“[A]s Mr. Riley candidly admitted in his deposition, he was ‘pushed’ further in his analysis at Troedel’s trial than at Hawkins’ trial. . . . [At the] evidentiary hearing held before this Court, one of the prosecutors testified that, at Troedel’s trial, after Mr. Riley had rendered his opinion which was contained in his written report, the prosecutor *pushed* to ‘see if more could have been gotten out of this witness.’”).

465. NRC I, *supra* note 24, at 149 (“Because of the potential power of DNA evidence, authorities must make funds available to pay for expert witnesses. . . .”).

466. President’s DNA Initiative, *supra* note 450.

467. 470 U.S. 68 (1985); see Paul C. Giannelli, *Ake v. Oklahoma: The Right to Expert Assistance in a Post-Daubert, Post-DNA World*, 89 Cornell L. Rev. 1305 (2004).

468. *Ake*, 470 U.S. at 74.

469. 18 U.S.C. § 3006(A).