Case Western Reserve Law Review

Volume 68 | Issue 3 Article 13

2018

Firearm-Mark Evidence: Looking Back and Looking Ahead

David H. Kaye

David H. Kaye[†]

FIREARM-MARK EVIDENCE: LOOKING BACK AND LOOKING AHEAD

Paul Giannelli has written—with clarity and conviction—on just about every type of scientific evidence commonly used in criminal cases. To celebrate his extraordinary contributions, this Article surveys the development of the law on one type of feature-matching evidence that repeatedly attracted Paul's attention. This summary reinforces and extends Paul's work on what this Article will call "firearm-mark evidence."[1] By inspecting toolmarks on bullets or spent cartridge cases, firearms examiners can supply valuable information on whether a particular gun fired the ammunition in question. But the limits on this information have not always been respected in court, and a growing number of opinions have tried to address this fact. Reviewing this development is significant not merely because the evidence is commonly employed in criminal cases, but also because of a recent, highly publicized[2] argument against its admission from some of the nation's

† Distinguished Professor of Law and Weiss Family Scholar, Penn State Law.

1. "Although this subject is popularly known as 'ballistics,' that term is not correct." 1 PAUL C. GIANNELLI ET AL., SCIENTIFIC EVIDENCE § 14.01, at 755 (5th ed. 2012).

2. *E.g.*, Alex Kozinski, *Rejecting Voodoo Science in the Courtroom*, WALL ST. J. (Sept. 19, 2016, 7:36 PM), https://www.wsj.com/articles/rejecting-voodoo-science-in-the-courtroom-1474328199 [https://perma.cc/4WAH-NBNW].

leading scientists and technologists[3] and because it can inform a pending effort to improve the federal rules as they apply to forensic-science identification evidence.[4]

As we shall see, the courts have moved from a position of skepticism of the ability of examiners to link bullets and other ammunition components to a particular gun to full-blown acceptance of claims of identification "to the exclusion of all other firearms."[5] Although challenges to firearm-mark evidence over the past decade or so have generated occasional restrictions on the degree of confidence that firearms experts can express in their source identifications, they have not altered the paradigm of supplying source conclusions instead of statements about the degree to which the evidence supports these conclusions.[6] After reviewing the stages in the judicial reception of firearm-mark evidence, this Article concludes by describing a more scientific, quantitative, evidence-based form of testimony that should supplant or augment the current experience-based decisions of skilled witnesses.

I. Rejection of Expert Source Attributions

For a time, courts did not admit testimony that items originated from a particular firearm. Some courts reasoned that jurors could make

3. EXEC. OFFICE OF THE PRESIDENT, PRESIDENT'S COUNSEL OF ADVISORS ON SCI. & TECH., REPORT TO THE PRESIDENT: FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016) [hereinafter 2016 PCAST REPORT].

4. David H. Kaye, *How* Daubert *and Its Progeny Have Failed Criminalistics Evidence and a Few Things the Judiciary Could Do About It*, 86 FORDHAM L. REV. 1639 (2018).

5. *E.g.*, *In re* Barrett, 840 F.3d 1223, 1238 (10th Cir. 2016) ("Ballistics expert Terrance Higgs tied the bullet fragment that killed Eales to Defendant's .223 Colt H Bar Sporter rifle, 'to the exclusion of all guns that are made or that will be made.'"); United States v. Law, 252 F.3d 1357, 2001 WL 422948, at *1 (5th Cir. 2001) ("[B]allistics expert testified that the cartridge recovered at the earlier robbery and the cartridge used in the Griffin carjacking were used in the same weapon 'to the exclusion of all other firearms in the world.'").

6. In this context, a source conclusion is a statement about the truth or probability of the hypothesis that a specific, known gun fired the bullet in question. Statements of support stop short of drawing a conclusion about the hypothesis. Instead, they describe the probability of the evidence—the extent to which the features of the items being compared are observed to correspond—under competing source hypotheses. *See* DAVID H. KAYE ET AL., THE NEW WIGMORE: A TREATISE ON EVIDENCE—EXPERT EVIDENCE ch. 14 (2d ed. 2011); David H. Kaye, *Hypothesis Testing in Law and Forensic Science: A Memorandum*, 130 HARV. L. REV. F. 127 (2017); *infra* Part VI.

the comparisons and draw their own conclusions. In *People v. Weber*,[7] for example, the trial court struck from the record an examiner's testimony "that in his opinion the two bullets taken from the bodies were fired from this pistol, leaving that as a question for the jury to determine by an inspection of the bullets themselves."[8] In this 1904 trial, the court did not question the expert's ability to discover toolmarks that could be probative of identity, but it saw no reason to believe that the expert would be better than lay jurors at drawing inferences from that information.[9] Other courts allowed such opinions, but not if they were stated as "facts."[10]

II. ACCEPTANCE OF EXPERT SOURCE ATTRIBUTIONS

With the recognition that the line between "opinions" and "facts" had little substance and with the demise of the rigid rule prohibiting "ultimate facts"—which were said to "invade the province of the jury"[11]—courts came to admit conclusive source attributions. Firearms examiners reasoned that

> It may be quite common for two or more prominent individual marks on bullets from two entirely different guns to match exactly, but the chance that there will be a correspondence of a great many of the individual characteristic marks on two bullets that came from different guns is so remote as to amount to a practical impossibility.[12]

By the 1950s, it was understood that:

7. 86 P. 671 (Cal. 1906).

8. *Id.* at 678.

9. The court explained that "the comparison of the . . . bullets . . . is not a matter of expert testimony, but one within the ordinary capacities of the average juror or citizen." *Id.*

10. E. LeFevre, Annotation, *Expert Evidence to Identify Gun from Which Bullet or Cartridge Was Fired*, 26 A.L.R. 2d 892, § 3, at 898 (1952). For example, in *State v. Martinez*, the state supreme court held that testimony that stated "positively that the evidence bullet (death bullet) was fired out of [defendant's] gun" was an instance of inadmissible "conclusions stated as facts and not as opinions." 198 P.2d 256, 260–61 (N.M. 1948).

11. *E.g.*, Grismore v. Consolidated Prods. Co., 5 N.W.2d 646, 655 (Iowa 1942), *overruling* State v. Steffen, 230 N.W. 536, 538 (Iowa 1930).

12. JULIAN S. HATCHER, TEXTBOOK OF FIREARMS INVESTIGATION, IDENTIFICATION AND EVIDENCE 287–88 (1st ed. 1935); *cf.* ALBERT S. OSBORN, QUESTIONED DOCUMENTS 227–30 (2d ed. 1929) (asserting that duplication of class and individual characteristics of handwriting can be "practically impossible" because the joint probability is a "negligible quantity").

> [T]he modern tendency of the courts [is] to allow the introduction of expert testimony to show that the bullet or cartridge found at the scene of a crime was fired from a particular gun, where it is definitely shown that the witness by whom the testimony is offered is, by experience and training, qualified to give an expert opinion on firearms and ammunition.[13]

Firearms and other types of examiners were known to testify that their judgments are not subject to any margin of error[14] and are scientific certainties.[15] Of course, expert testimony was not required to be so extreme; testimony that a bullet merely could or might have come from a particular firearm also was admissible.[16]

III. HEIGHTENED SCRUTINY FOLLOWING *DAUBERT*

Beginning in the 1990s, scientists and lawyers began to question the theories of individualization and discernible uniqueness of firearms toolmarks. They asked how examiners, operating without standards explicitly defining what degree of similarity in a set of features warrants a source attribution, could *know*—in the sense described in *Daubert v. Merrell Dow Pharmaceuticals*[17]—that a given gun fired the recovered items. A series of challenges to the admissibility of source attributions by firearms examiners ensued, and professional examiners responded with an "Admissibility Resource Kit" to "assist firearm examiners in better preparing for evidence admissibility hearings that began to greatly proliferate in 2002."[18]

---

13. LeFevre, *supra* note 10, § 5, at 901.

14. Watkins v. Commonwealth, 331 S.E.2d 422, 434 (Va. 1985). The Virginia Supreme Court saw no problem with "[t]his positive statement" which "merely affects the weight of his testimony" and "does not necessarily invalidate or even weaken the results of his ballistics testing." *Id.*

15. United States v. Natson, 469 F. Supp. 2d 1253, 1261 (M.D. Ga. 2007) (noting that FBI supervisory special agent Paul Tangren "opined that he held this opinion to a 100% degree of certainty.").

16. GIANNELLI ET AL., *supra* note 1, § 14.06[a], at 773; Jay M. Zitter, Annotation, *Admissibility of Testimony That Bullet Could or Might Have Come from Particular Gun*, 31 A.L.R. 4th 486, 487 (1984).

17. 509 U.S. 579 (1993). *Daubert* interpreted the phrase "scientific knowledge" in Federal Rule of Evidence 702 to mean "derived by the scientific method . . . supported by appropriate validation—i.e., 'good grounds,' based on what is known." *Id.* at 590. An untold number of cases have attempted to apply these generalities. *See, e.g.*, GIANNELLI ET AL., *supra* note 1; KAYE ET AL., *supra* note 6, § 7.3.

18. *SWGGUN Admissibility Resource Kit (ARK)*, ASS'N FIREARM & TOOL MARK EXAM'RS [hereinafter *Admissibility Resource Kit]*, https://afte.org/resources/swggun-ark [https://perma.cc/F78P-KRPF] (last visited Feb.

Initially, the courts were unfazed by the post-*Daubert* skepticism about what they comfortably knew as "a recognized method of ballistics testing"[19] that "has been accepted in criminal cases for many years."[20] But then a number of federal district courts expressed misgivings about holistic judgments of "sufficient agreement of individual characteristics."[21] No court excluded all evidence of similarities, but several struggled to find ways to allow examiners to assist the jury without testifying that cartridge components definitely came from the known firearm or that nothing else was scientifically or practically possible. The first such case during this period was *United States v. Green.*[22] In a summary of cases in this period, Paul called the opinion, written by U.S. District Judge Nancy Gertner, "riveting."[23] It restricted the firearms examiner to testifying about the matching features—a reversion to the *Weber* era.[24] The expert admitted that in applying the Association of Firearms and Toolmark Examiners' ("AFTM") theory

11, 2018); *cf.* Kirsten Jackson, *The* Daubert *Era*, *in* SCIENTIFIC EXAMINATION OF QUESTIONED DOCUMENTS 37, 41 (Jan Seaman Kelly & Brian S. Lindblom eds., 2d ed. 2006) (attributing success in rebuffing "over 30 *Daubert* challenges" to handwriting identification to "the *Daubert* Group" formed by the American Board of Forensic Document Examiners).

19. United States v. Hicks, 389 F.3d 514, 526 (5th Cir. 2004) ("[T]he matching of spent shell casings to the weapon that fired them has been a recognized method of ballistics testing in this circuit for decades.").

20. United States v. Foster, 300 F. Supp. 2d 375, 376 n.1, 377 (D. Md. 2004) (reasoning that "the 'human ability to recognize a similar pattern and distinguish between dissimilar patterns' makes identification possible"). Some courts frankly declined to require compliance with all the *Daubert* factors. *E.g.*, United States v. Santiago, 199 F. Supp. 2d 101, 112 (S.D.N.Y. 2002) (stating that acceptance "in the community of forensics experts" can substitute for acceptance in "a scientific community"). For more strategies used to avoid the strictures of *Daubert* for criminalistics identification evidence, see Kaye, *supra* note 4.

21. *Admissibility Resource Kit*, *supra* note 18; *cf. AFTE Theory of Identification as It Relates to Toolmarks*, ASS'N FIREARM & TOOL MARK EXAM'RS, https://afte.org/about-us/what-is-afte/afte-theory-of-identification [https://perma.cc/3EQF-7JQ9] (last visited Feb. 11, 2018) ("[S]ufficient agreement" for "subjective" "individualization/identification" occurs "when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool.").

22. 405 F. Supp. 2d 104 (D. Mass. 2005).

23. Paul C. Giannelli, *Ballistics Evidence Under Fire*, CRIM. JUST., Winter 2011.

24. *See supra* Part I.

of sufficiency,[25] "it's just your opinion? You determine which marks you're going to pay attention to and which ones you're not, correct?"[26] The court found the examiner's assurance "that this match could be made 'to the exclusion of every other firearm in the world'" to be "extraordinary, particularly given [the] data and methods."[27] In view of the method's subjectivity, potential for bias, and lack of data on error rates, the district court perceived "no accurate way of evaluating the testimony."[28]

No other modern, published opinion has confined the examiner to reporting on similarities and differences in the toolmarks.[29] Instead, a few concerned courts focused on how firmly an examiner could characterize source attributions. In *United States v. Monteiro*,[30] another federal district judge in the same district adopted the more lenient rule that "the expert may testify that the cartridge cases were fired from a particular firearm to a reasonable degree of ballistic certainty. However, the expert may not testify that there is a match to an exact statistical certainty."[31]

Seeking a less opaque formulation, District Judge Jed Rakoff in *United States v. Glynn*[32] excluded testimony of "a reasonable degree of ballistic certainty"[33] in favor of a weaker statement of "more likely than not."[34] This conclusion-lite testimony, along with other evidence in the case, still led to a conviction and life sentence.[35]

25. *See supra* note 20.

26. *Green*, 405 F. Supp. 2d at 112 n.15 (citation omitted).

27. *Id.* at 107 (citation omitted).

28. *Id.* at 121 (footnote omitted).

29. For discussion of unadorned "features only testimony" and single-stage "'not excluded' or 'match'" testimony for scientific identification evidence, see KAYE ET AL., *supra* note 6, §§ 15.3–15.4.

30. 407 F. Supp. 2d 351 (D. Mass. 2006).

31. *Id.* at 375.

32. 578 F. Supp. 2d 567 (S.D.N.Y. 2008).

33. *Id.* at 574.

34. *Id.* at 575. GIANNELLI ET AL., *supra* note 1, § 14.06[b], at 776–777, suggests that *Monteiro* used the same standard. However, the only use of the phrase is in a citation to a case involving bite-mark evidence as one illustration of the type of testimony that would fall short of the "100 percent sure" formulation that the court excluded in favor of "reasonable degree of ballistic certainty." *Monteiro*, 407 F. Supp. 2d at 372.

35. Press Release, U.S. Attorney for the S. Dist. of N.Y., Bloods Gang Member Sentenced to Life in Prison for Ordering a Drug-Related Murder in 2000 (Jan. 28, 2009), https://www.justice.gov/archive/usao/nys/pressreleases/January09/glynnsentencingpr.pdf [https://perma.cc/C78B-AW8J].

The *Glynn* court denied that firearms source attributions "could . . . be called 'science,'" [36] because when asked "what constitutes 'sufficient agreement' between two pieces of ballistic evidence to declare a match, [the government's expert] admitted that the assessment is subjective, in that 'it is an opinion of mine and whether or not someone else would agree with it is up to that individual.'"[37]

The *Glynn* court may have been influenced by a report of a committee of the National Academy of Sciences.[38] This NAS committee was formed to assess the feasibility of creating a computer-searchable national database "that would house images of firings of all newly manufactured and imported firearms . . . as an aid to criminal investigations." [39] Although the committee was concerned with digital imaging and pattern-recognition technology, it began with an inquiry into the logic of traditional firearm-mark analysis.[40] It reported that "[t]he validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated."[41] Moreover, the committee approved of opinions that "refused to accept 'exclusion of all other firearms' arguments"[42] and disapproved of the practice of "overreach[ing] to make extreme probability statements."[43]

36. 578 F. Supp. 2d at 570.

37. *Id.* at 571 (citation omitted). Thus, the court found that the AFTE "standard defining when an examiner should declare a match—namely, 'sufficient agreement'—is inherently vague." *Id.* at 572.

38. *Id.*

39. NAT'L RESEARCH COUNCIL COMM. TO ASSESS THE FEASIBILITY, ACCURACY, AND TECH. CAPABILITY OF A NAT'L BALLISTICS DATABASE, BALLISTIC IMAGING 1–2 (Daniel L. Cork et al. eds., 2008) [hereinafter 2008 REPORT]. The committee concluded that such a database would not be advisable, but recommended enhancements to the existing National Integrated Ballistic Information Network. *Id.* at 4–6.

40. *Id.* at 3 ("Underlying the specific tasks with which the committee was charged is the question of whether firearms-related toolmarks are unique: that is, whether a particular set of toolmarks can be shown to come from one weapon to the exclusion of all others. Very early in its work the committee found that this question cannot now be definitively answered.").

41. *Id.* at 3, 81.

42. *Id.* at 82.

43. *Id.* at 85. The AFTE disagreed. It maintained, as it always has, that examiners can and do achieve practical scientific certainty. AFTE Comm. for the Advancement of the Sci. of Firearm & Toolmark Identification, *The Response of the Association of Firearm and Tool Mark Examiners to the National Academy of Sciences 2008 Report Assessing the Feasibility, Accuracy, and Technical Capability of a National Ballistics Database*, AFTE J., Summer 2008, at 234, 242 [hereinafter AFTE Comm. Response]. However, the AFTE's definition of "practical certainty" for "a scientific

IV. HEIGHTENED SCRUTINY FOLLOWING THE 2009 NAS REPORT

Soon after the 2008 NAS report, a larger NAS Committee on Identifying the Needs of the Forensic Sciences Community observed that "[m]uch forensic evidence—including, for example, bite marks and firearm and toolmark identifications—is introduced in criminal trials without any meaningful scientific validation . . . ."[44] The committee reiterated some of the statements from the 2008 report,[45] emphasized the need for valid estimates of the uncertainties in forensic-science identification methods generally,[46] and pointed to a way to express the probative value of the associations without drawing a source conclusion.[47]

Neither the 2008 nor the 2009 NAS report made recommendations on admissibility of evidence, for that was not part of their charge.[48] Practitioners and prosecutors proposed that this meant that the reports

conclusion" is surprisingly weak. It means only that "an examiner . . . believes the conclusion to be true and accurate; . . . has rational grounds for [the belief]; and acknowledges that, in the abstract, it is not possible to achieve absolute certainty for results flowing from a scientific theory or technique"; *cf.* John E. Murdock et al., *The Development and Application of Random Match Probabilities to Firearm and Toolmark Identification*, 62 J. FORENSIC SCI. 619, 624 (2017) ("Absolute certainty opinions may have been adopted in the past, but this type of position has been retired for some time and no longer represents the consensus thinking of the firearm and toolmark community. . . . [O]ur everyday lives are predicated upon practical certainty. There is a practical certainty that our car will start in the morning (assuming it is in good mechanical condition), or that our (normally obedient) dog will come when called.").

44. NAT'L RESEARCH COUNCIL, COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIS. CMTY., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 107–08 (2009) [hereinafter NAT'L RESEARCH COUNCIL REPORT] (citations omitted).

45. *Id.* at 154.

46. *Id.* at 184.

47. The committee remarked that "[p]ublications such as Evett et al., Aitken and Taroni, and Evett provide the essential building blocks for the proper assessment and communication of forensic findings." *Id.* at 186 (citations omitted). Such publications advocate strength-of-evidence statements rather than source conclusions.

48. Indeed, the 2008 committee cautioned that "*the proposal for this study explicitly precluded the committee from assessing the admissibility of forensic firearms evidence in court*, either generally or in specific regard to testimony on ballistic imaging comparisons." 2008 REPORT, *supra* note 39, at 20. In the next breath, the committee added the following: "We note, however, that high-subjectivity branches of forensic science are now confronting growing skepticism with regard to discernible uniqueness as a result of a number of legal and scientific studies." *Id.*

should or could not be taken as undermining the admissibility of traditional highly judgmental pattern-matching identifications.[49] However, the committees' reviews of the literature clearly lent credence to the questions about the routine admission of categorical source attributions based on firearm-marks.[50] In five prominent published opinions, courts

49. *E.g.*, AFTE Comm. Response, *supra* note 43, at 241–42; Government's Opposition to Defendant's Motion to Exclude Expert Testimony Concerning Latent Fingerprint Evidence at 3, United States of America v. Faison, No. 2008-CF2-16636 (D.C. Super. Ct. Feb. 19, 2010), *quoted in* Harry T. Edwards, *The National Academy of Sciences Report on Forensic Sciences: What It Means for the Bench and Bar*, 51 JURIMETRICS J. 1, 5–6 (2010) (describing this argument as "utterly absurd").

50. For example, in describing the scientific basis of "forensic science fields like firearms examination," the 2008 report quoted with approval an article by two forensic scientists stating that "[f]orensic individualization sciences that lack actual data, which is most of them, . . . simply . . . assume the conclusion of a miniscule probability of a coincidental match . . . ." 2008 REPORT, *supra* note 39, at 54–55 (quoting John I. Thornton & Joseph L. Peterson, *The General Assumptions and Rationale of Forensic Identification*, *in* 3 DAVID L. FAIGMAN, DAVID H. KAYE, MICHAEL J. SAKS, & JOSEPH SANDERS, MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 24-7.2, at 169 (2002)). Apparently recognizing the threat of such assessments, AFTE complained that the committees' literature reviews were shallow. In response to the 2008 Report, it wrote that "the committee lacked the expertise and information necessary for the in-depth study that would be required to offer substantive statements with regard to these fundamental issues of firearm and toolmark identification." AFTE Comm. Response, *supra* note 43, at 243. Likewise, it wrote that "the [2009] NAS committee in effect chose to ignore extensive research supporting the scientific underpinnings of the identification of firearm and toolmark evidence." AFTE Comm. for the Advancement of the Sci. of Firearm & Toolmark Identification, *The Response of the Association of Firearms and Tool Mark Examiners to the February 2009 National Academy of Science Report "Strengthening Forensic Science in the United States: A Path Forward,"* AFTE J., Summer 2009, at 204, 206. According to AFTE, "years of empirical research . . . conclusively show[] that sufficient individuality is often present on tool (firearm tools or non-firearm tools) working surfaces to permit a trained examiner to conclude that a toolmark was made by a certain tool and that there is no credible possibility that it was made by any other tool working surface." AFTE Comm. Response, *supra* note 43, at 242. After all, "[t]he principles and techniques utilized in forensic firearms identification have been *used internationally* for nearly a century *by the relevant forensic science community* to both identify and exclude specific firearms as the source of fired bullets and cartridge cases." *Id.* at 237 (emphasis added). Prosecutors too sought to blunt the implications of the skeptical statements about the limited validation of the premises of the traditional theory of firearm-mark identification with an affidavit from the chairman of the NAS committee that wrote the 2008 Report. Affidavit of John E. Rolph at 1–3, United States v. Edwards, No. F-516-01 (D.C. Super. Ct. May 23, 2008). Yet, the affidavit merely collects excerpts from the report itself and ends with one that could be read as supporting admissibility under certain conditions. For another affidavit from

cited the NAS reports and the opinions cited in in Part III of this Article to limit such testimony. First, the district court in *United States v. Taylor*[51] deemed the AFTE theory of sufficiency "circular."[52] It reiterated the assessment of the 2009 NAS committee that "a fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process. . . . AFTE has adopted a theory of identification, but it does not provide a specific protocol."[53] To cope with the absence of controlling standards for making source attributions, the court held that the expert "will not be permitted to testify that his methodology allows him to reach this conclusion as a matter of scientific certainty [or] . . . that there is a match to the exclusion, either practical or absolute, of all other guns."[54] Instead, "[h]e 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."[55]

Second, *United States v. Willock*[56] provides the most extensive judicial analysis of firearms testimony to date. It observes that "toolmark analysis guidance provided by the AFTE lacks specificity because it allows an examiner to identify a match based on 'sufficient agreement,' which the AFTE defines using the undefined terms 'exceeds the best agreement' and 'consistent with.'"[57] Based on "reading . . . the many published studies, journal articles, and cases," Magistrate Judge Paul Grimm characterized "the AFTE theory . . . that once 'sufficient

a committee member contending that NAS "has questioned the validity of these fundamental assumptions of uniqueness and reproducibility," see Declaration of Alicia Carriquiry, PhD. In Support of Motion in Limine to Exclude Firearms Examiner's Opinion at 5, People v. Knight, No. LA067366 (Cal. Super. Ct. Apr. 2012). The use of affidavits of one or two committee members to give their personal views on what the words that the committee as a whole agreed upon is ill-advised. It resembles asking individual members of Congress to provide their *post hoc* thoughts on what a committee report on legislation, or the statute itself, really meant.

51. 663 F. Supp. 2d 1170 (D.N.M. 2009).

52. *Id.* at 1177.

53. *Id.* at 1178 (quoting NAT'L RESEARCH COUNCIL REPORT, *supra* note 44, at 155).

54. *Id.* at 1180.

55. *Id.*

56. 696 F. Supp. 2d 536 (D. Md. 2010).

57. *Id.* at 566 (quoting NATIONAL RESEARCH COUNCIL REPORT, *supra* note 44, at 155).

agreement' [establishes] a practical impossibility" as "astonishing[]."[58] The district court ordered "[t]hat [the expert] not be allowed to opine that it is a 'practical impossibility' for any other firearm to have fired the cartridges . . . [and that he] only be permitted to state his opinions and bases without any characterization as to degree of certainty."[59]

Third, in *Commonwealth v. Pytou Heang*,[60] the Massachusetts Supreme Judicial Court enumerated difficulties with the AFTE theory of sufficiency and practical impossibility. It settled on "reasonable degree of ballistic certainty" as an acceptable indication of the limits of an opinion, and cautioned that "[p]hrases that could give the jury an impression of greater certainty, such as 'practical impossibility' and 'absolute certainty' should be avoided."[61] Likewise, it ruled that "'reasonable degree of scientific certainty' should . . . be avoided because it suggests that forensic ballistics is a science, where it is clearly as much an art as a science."[62]

Fourth, the district court in *United States v. Ashburn*[63]—while declining to go as far as *Green* and *Glynn* in circumscribing source opinions—relied on the 2009 NAS Report and the criticisms of the AFTE sufficiency theory in the opinions discussed above to preclude "this expert witness from testifying that he is 'certain' or '100%' sure . . . [or] that a match he identified is to 'the exclusion of all other firearms in the world,' or that there is a 'practical impossibility' that any other gun could have fired the recovered materials."[64] It limited the expert "to stating that his conclusions were reached to a 'reasonable degree of ballistics certainty' or a 'reasonable degree of certainty in the ballistics field.'"[65]

58. *Id.* at 572 (quoting Ronald G. Nichols, *Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges*, 52 J. FORENSIC SCI. 586, 590 (2007)).

59. *Id.* at 581–82.

60. 942 N.E.2d 927 (Mass. 2011).

61. *Id.* at 946.

62. *Id.*; *cf.* United States v. Cazares, 788 F.3d 956, 989 (9th Cir. 2015) (distinguishing between "scientific certainty" and "a reasonable degree of certainty in the ballistics field," holding that the latter expression "is the proper expert characterization of toolmark identification," and failing to consider whether a report of "practical impossibility" would be admissible).

63. 88 F. Supp. 3d 239 (E.D.N.Y. 2015).

64. *Id.* at 249.

65. *Id.*

Finally, in *Gardner v. United States*,[66] the District of Columbia Court of Appeals, without mentioning *Willock*, wrote that it was error to admit an examiner's "unqualified opinion."[67] The court cited "questions about pattern matching generally, and bullet pattern matching specifically, [that] surfaced in the scientific community."[68] Although the opinion condemned "absolute or 100% certainty," it did not specify the qualifications an examiner would have to place on source attributions, and it did not discuss the AFTE theory of sufficiency for "practical impossibility."[69]

To be clear, the cases collected here are exceptions to the normal, uncritical acceptance of firearm-mark testimony. And during this same period, other courts, in less detailed opinions, imposed no limitations on source attributions.[70] In all, the modern opinions on firearms source attribution uniformly hold that the similarities in the features can be presented—just as the earliest opinions on the subject did—and all but one allow an expert to provide some opinion on the source hypothesis. But what kind of an opinion that should be is being probed with increasing frequency. Although the still small number of critical cases are all over the map on how such opinions can or should be presented, this developing line of authority does seem to reflect a growing judicial sense of unease about the AFTE theory of personal sufficiency and practical impossibility, and no firm support for the theory is apparent in the legal commentary. To the contrary, legal commentators tend to criticize the modern opinions for not excluding all conclusions based on current

66. 140 A.3d 1172 (D.C. 2016).

67. *Id.* at 1184.

68. *Id.* at 1183.

69. *Id.* at 1184.

70. *E.g.*, United States v. Casey, 928 F. Supp. 2d 397, 399–400 (D.P.R. 2013) (stating that although "defendant challenges [the] conclusion that [the examiner] is 100% certain," the court "remains faithful to the long-standing tradition of allowing the unfettered testimony of qualified ballistics experts"); United States v. Natson, 469 F. Supp. 2d 1253, 1261–62 (M.D. Ga. 2007) (permitting forensic ballistics expert to offer an opinion of a match "to a 100% degree of certainty"); State v. Davidson, 509 S.W.3d 156, 205 (Tenn. 2016) ("It's like a fingerprint.").

methods for comparisons[71] or for allowing "extremely misleading" phrases for a degree of certitude in a source attribution.[72]

V. THE 2016 PCAST REPORT

A third report from scientists outside of the firearms and toolmarks community generated even more consternation within that community and among law enforcement officials.[73] Late in 2016, the President's Council of Advisors on Science and Technology ("PCAST") released a report on "ensuring scientific validity of feature-comparison methods."[74] Like the two NAS reports, the PCAST report questions the AFTE theory of unstructured firearm-mark identification to a practical certainty. Indeed, it dismisses it as "clearly not a scientific theory," but rather "a claim that examiners applying a subjective approach can

71. *E.g.*, 4 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 34:5–6 (2016–2017 ed.). This treatise refers to "cases like *Green*, *Glynn*, and *Willock*" as "partial and somewhat unsatisfying" and "a mere band-aid, requiring experts to slightly soften the language in which they express their conclusions, but not requiring any more significant modifications, nor any concrete empirical evidence regarding error rates, nor objective metrics to guide comparisons." *Id.* § 34:5, at 893; *see also* KAYE ET AL., *supra* note 6, § 15.2.4, at 685 (describing the *Monteiro* line of cases as allowing "the expert [to] give a looser opinion intended to connote that even if there is some chance of a matching weapon somewhere in the world, the bullet very likely passed through the barrel of the gun in the case at bar" and observing that "[w]hether even this weaker statement of local individualization satisfies *Daubert* and *Kumho Tire* is open to serious question").

72. GIANNELLI ET AL., *supra* note 1, § 14.06[c], at 780; *cf.* KAYE ET AL., *supra* note 6, § 15.2.4, at 685 ("'[T]o a reasonable degree of scientific certainty' adds nothing meaningful to the opinion . . . ."); *id.* § 15.5, at 698 ("Unless the source probability is demonstrably very close to one, so that a source attribution is defensible, nonnumerical expressions of source probability do not seem promising.").

73. For discussion of early reactions of the forensic science establishment, see Adam B. Shniderman, *Prosecutors Respond to Calls for Forensic Science Reform: More Sharks in Dirty Water*, 126 YALE L.J. F. 348 (2017); David H. Kaye, *The National District Attorneys Association's Slam: PCAST "Usurps the Constitutional Role of the Courts"*, FORENSIC SCI., STAT. & L. (Sept. 5, 2016), http://for-sci-law.blogspot.com/2016/09/the-national-district-attorneys.html [https://perma.cc/6FTT-GQE2]; David H. Kaye, *The PCAST Report and Argumentum Ad Hominem,* FORENSIC SCI., STAT. & L. (Sept. 24, 2016), http://for-sci-law.blogspot.com/2016/09/the-pcast-report-and-argumentum-ad.html [https://perma.cc/9BQ8-FWQN].

74. 2016 PCAST REPORT, *supra* note 3.

accurately individualize the origin of a toolmark" based on a "stated method" that "is circular."[75]

A. Validity of Traditional Firearm-Mark Analysis

The report finds that, whatever the theory behind firearm-mark analysis may be, the AFTE procedure has yet to be validated. Finding Six is blunt: "PCAST finds that firearms analysis currently falls short of the criteria for foundational validity, because there is only a single appropriately designed study to measure validity and estimate reliability. The scientific criteria for foundational validity require more than one such study, to demonstrate reproducibility."[76]

---

75. *Id.* at 60. In a reply to PCAST, the Firearms and Toolmark Subcommittee of the Organization of Scientific Area Committees for Forensic Science argued that the notion of sufficiency as the criterion for individualization is not circular because:

> The sufficient agreement threshold is exhibited when the amount of agreement is greater than best known non-matches established by the community and conveyed to each examiner through a lengthy and extensive training program. That is, it is not an arbitrary point. In fact, by definition, no non-matches can ever have more similarity than the sufficient agreement point.

ORG. OF SCI. AREA COMMS. (OSAC), FIREARMS AND TOOLMARKS SUBCOMM., RESPONSE TO THE PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY (PCAST) CALL FOR ADDITIONAL REFERENCES REGARDING ITS REPORT "FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS" 9 (2016) [hereinafter OSAC RESPONSE], https://www.theiai.org/president/20161214_FATM_Response_to_PCAST.pdf [https://perma.cc/4AM5-32FX]; *accord*, ASS'N FIREARM & TOOLMARK EXAM'RS, RESPONSE TO PCAST REPORT ON FORENSIC SCIENCE (2016), https://afte.org/uploads/documents/AFTE-PCAST-Response.pdf [https://perma.cc/5X52-9ZUC]. The idea is that examiners draw on a kind of internal database—an overall sense of the similarity of some set of the most closely matching pairs of items from different sources that they encountered when they were trained or in exercises since then. They compare their memory of the similarities in different-source specimens to the observed similarities in the current case. If the current pair is outside the remembered range for non-mates, they believe that it is logically impossible for the current pair to have originated from the same source ("by definition," that cannot occur). It seems doubtful that most courts would agree that this articulation provides the "specificity" required to avoid the kind of "circularity" or "inherent vagueness" that troubled the courts in *Taylor*, *Willock*, and *Glynn*.

76. 2016 PCAST REPORT, *supra* note 3, at 112; *see also id.* at 111 ("The scientific criteria for foundational validity require appropriately designed studies by *more than one group* to ensure reproducibility. Because there has been only a single appropriately designed study, the current evidence falls short of the scientific criteria for foundational validity."). The response from the OSAC subcommittee maintains that other types of studies supply ample proof of validity. OSAC RESPONSE, *supra* note 75, at 2–7. In an addendum

This damning conclusion follows from the specific criteria that PCAST adopted for establishing what it called "foundational validity."[77] Finding One of the report explains that:

> To establish foundational validity for a forensic feature-comparison method, the following elements are required: (a) a reproducible and consistent procedure for (i) identifying features in evidence samples; (ii) comparing the features in two samples; and (iii) determining, based on the similarity between the features in two sets of features, whether the samples should be declared to be likely to come from the same source ("matching rule"); and (b) empirical estimates, from appropriately designed studies from multiple groups, that establish (i) the method's false positive rate—that is, the probability it declares a proposed identification between samples that actually come from different sources and (ii) the method's sensitivity—that is, the probability it declares a proposed identification between samples that actually come from the same source.[78]

Among other things, the scientific validation studies "should be conducted so that the examinees have no information about the correct answer."[79] Furthermore, for source conclusions that are not the product of a standardized, step-by-step procedure that involves "little or no judgment,"[80] PCAST insists on one, and apparently only one, approach to establishing foundational validity—"the method must be evaluated as if it were a 'black box' in the examiner's head"[81] via "black-box studies that measure how often many examiners reach accurate conclusions

to the 2016 report, PCAST reiterated that the designs of most of the other studies are too flawed to permit them to be relied on to establish validity. PCAST, AN ADDENDUM TO THE PCAST REPORT ON FORENSIC SCIENCE IN CRIMINAL COURTS 7 (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf [https://perma.cc/EH8W-FQUD] ("These studies do not provide useful information about the actual reliability of firearms analysis."). It conceded that two additional studies, although still flawed, merited some consideration. *Id.*

77. "Foundational validity" is not a standard phrase in metrology and statistics. "Validity" as PCAST defined it is discussed in KAYE ET AL., *supra* note 6, § 15.7.5(c) (Cum. Supp. 2017), and Kaye, *supra* note 4, at 1653–54.

78. 2016 PCAST REPORT, *supra* note 3, at 65.

79. *Id.* at 66.

80. *Id.* at 5 n.3.

81. *Id.* at 5.

across many feature-comparison problems involving samples representative of the intended use."[82]

By applying the *no*-information-about-the-correct-answer criterion, PCAST narrowed the number of "appropriately designed" studies to one unpublished experiment.[83] The "Ames Laboratory study" was funded by the Department of Energy and reported in 2014.[84] The 218 examiners who elected to participate "made . . . 15 comparisons of 3 knowns to 1 questioned cartridge case. For all participants, 5 of the sets were from known same-source firearms [known to the researchers but not the firearms examiners], and 10 of the sets were from known different-source firearms."[85] Ignoring "inconclusive" comparisons, the performance of the examiners is shown in Table 1.

Table 1. Associations of Cartridge Cases to Handguns in the Ames Laboratory Performance Study (Baldwin 2014).

	~S	+S	
–E	1421	4	1425
+E	22	1075	1097
	1443	1079	

–E is a negative finding (the examiner decided there was no association).

82. *Id.* at 66. For both objective and subjective methods, "[t]he studies must (a) demonstrate that the method is repeatable and reproducible and (b) provide valid estimates of the method's accuracy (that is, how often the method reaches an incorrect conclusion) that indicate the method is appropriate to the intended application." *Id.* at 5. "Repeatable" and "reproducible" are terms of art in metrology.

> Repeatability describes the agreement within sets of measurements . . . where the same person uses the same equipment in the same way under the same conditions (including place and, as far as possible, time). Reproducibility . . . describes the agreement within a set of measurements . . . where different people, equipment, methods or conditions are involved.

MIKE GOLDSMITH, NAT'L PHYSICAL LAB., NAT'L MEASUREMENT SYS., GOOD PRACTICE GUIDE NO. 118, A BEGINNER'S GUIDE TO MEASUREMENT 21 (2010), http://www.npl.co.uk/publications/a-beginners-guide-to-measurement. [https://perma.cc/KT3M-B6LX].

83. 2016 PCAST REPORT, *supra* note 3, at 111.

84. DAVID P. BALDWIN ET AL., AMES LABORATORY, DEP'T OF ENERGY, A STUDY OF FALSE-POSITIVE AND FALSE-NEGATIVE ERROR RATES IN CARTRIDGE CASE COMPARISONS, TECHNICAL REPORT #IS-5207 (2014), https://afte.org/uploads/documents/swggun-false-postive-false-negative-usdoe.pdf [https://perma.cc/4VWZ-CPHK].

85. *Id.* at 10.

+E is a positive finding (the examiner decided there was an association). ~S indicates that the cartridges came from bullets fired by a different gun. +S indicates that the cartridges came from bullets fired by the same gun.

There were twenty-two positive findings among the 1443 comparisons for different sources, for an observed false-positive rate of 22/1443 = 1.52%.[86] Taken at face value, these results are encouraging. On average, examiners displayed high levels of accuracy, both for cartridge cases from the same gun (better than 99 percent specificity) and from different guns (better than 98 percent sensitivity). Firearms examiners are not reaching all these correct conclusions by chance. In addition, these figures apply to the classifications made by single examiners in isolation—assuming that all the participants completed the exercises by themselves. Having a second, independent examination and then reconciling any differences in the outcomes before reporting an association or exclusion should reduce the rates of error.

Even so, an examination of further details of the Ames study reinforces PCAST's doubts about relying on this one study to conclude that a wide cross-section of examiners can achieve high accuracy rates. To begin with, researchers enrolled 284 volunteer examiners in the study by sending out emails and announcements in newsletters.[87] Using volunteers often biases the results of an experiment.[88] Second, one-third of the volunteers did not submit answers,[89] so nonresponse bias is a further concern. Third, the volunteers who completed the tasks were told that that they were being tested to "benefit society by providing a better statistical evaluation of this common and important forensic discipline that will strengthen the legal system in its understanding of the value of firearms comparisons."[90] Finally, only one type of firearm

86. *Id.* at 15–17. The 95 percent confidence interval is 0.96 percent to 2.30 percent. Conversely, the observed true-positive rate (also called specificity) is 98.48 percent. The 95 percent confidence interval is 97.70 percent to 99.04 percent.

87. *Id.* at 8.

88. *See e.g.*, P. F. Pinsky et al., *Evidence of a Healthy Volunteer Effect in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial*, 165 AM. J. EPIDEMIOLOGY 874 (2007).

89. BALDWIN ET AL., *supra* note 84, at 8–9.

90. *Id.* at 25. On the one hand, they may have been motivated to perform exceptionally well because they wanted to show that their work is valuable. On the other hand, they may have been less motivated by the knowledge that it was just an experiment rather than a part of a criminal investigation

and ammunition was used,[91] and only impressions on cartridge cases were considered.

As this example suggests, a robust set of studies—with different selection methods and conditions—is required to establish validity across an entire domain.[92] But there are studies with other firearms that indicate that examiners can discern the matching item out of a set when they know that the set contains a cartridge case or bullet filed by the test gun. The 2016 report dismisses these as of no value in establishing validity because source attribution in this "closed set" situation does not lend itself to meaningful estimates of error rates and is much easier than making source attributions when the examiner does not know whether a bullet in the test set came from the gun.[93] The very small error rates reported from such studies thus may grossly exaggerate accuracy, but they still lend some support to the claim that the expertise demonstrated in the Ames study extends beyond the limited circumstances of that study.

Consequently, despite PCAST's concerted effort to supply definitive criteria for judicial findings of the requisite degree of scientific validity to admit the conclusions of subjective interpretations of perceived features,[94] courts could continue to find that a sufficient scientific foundation for bullet-mark evidence exists even though the PCAST scientists did not. The report convincingly contends that "[n]othing—not training, personal experience nor professional

---

and that no individual's mistakes would be revealed to laboratory management.

91. The experimenters selected the inexpensive Ruger SR9 semiautomatic 9-mm Lugar centerfire pistol. *Id.* at 5, 9. All the guns were new. The ammunition came from two lots made by one manufacturer. *Id.* at 9.

92. *See* HANS ZEISEL & DAVID KAYE, PROVE IT WITH FIGURES: EMPIRICAL METHODS IN LAW AND LITIGATION 69 (1997) ("Consistent findings across different studies of the same type with different groups also are valuable, since they reduce the chance that the initial observations are due to a peculiarity in one group of subjects.").

93. Once an examiner picks the one true match, all the declarations of nonmatches are automatically correct. Experiments with other "set-to-set" designs have less dramatic internal dependencies but still fail to meet PCAST's strict criterion for being informative.

94. *See, e.g.*, 2016 PCAST REPORT, *supra* note 3, at 4 ("[L]egal standards and scientific standards intersect. Judges' decisions about the admissibility of scientific evidence rest solely on *legal* standards . . . . But, these decisions require making determinations about scientific validity. It is the proper province of the scientific community to provide guidance concerning scientific standards for scientific validity, and it is on those scientific standards that PCAST focuses here."); *id.* at 5 ("[Foundational validity] is the *scientific* concept we mean to correspond to the *legal* requirement, in Rule 702(c), of 'reliable principles and methods.'").

practices—can substitute for adequate empirical demonstration of accuracy."[95] Nonetheless, there is still room to debate the threshold for an "adequate empirical demonstration."[96]

B. Error-Rates for Firearm-Mark Analysis

Apparently recognizing that its criteria for an adequate empirical foundation might be disputed, the PCAST report hedges its bet. The report acknowledges that "[w]hether firearms analysis should be deemed admissible based on current evidence is a decision that belongs to the courts,"[97] but urges that any courts that reject its pronouncements on scientific validity admit source attributions only when accompanied by quantitative estimates of the false-positive error rate as inferred from rigorous performance studies.[98]

---

95. *Id.* at 4.

96. *Id.* Finding 6 concludes:

> If firearms analysis is allowed in court, the scientific criteria for validity as applied should be understood to require clearly reporting the error rates seen in appropriately designed black-box studies (estimated at 1 in 66, with a 95 percent confidence limit of 1 in 46, in the one such study to date).

Id. at 112. In the Addendum, PCAST continued to insist that "[f]rom a scientific standpoint, scientific validity should require at least two properly designed studies to ensure reproducibility." Addendum, *supra* note 76, at 7. But it conceded that there was some useful information in two other studies. It wrote that "[t]he issue for judges is whether one properly designed study, together with ancillary evidence from imperfect studies, adequately satisfies the legal criteria for scientific validity." *Id.* at 7–8. Firearms examiners maintain that many other studies noted but deemed inappropriate in the 2016 report comprise important evidence. OSAC Subcommittee Response, *supra* note 75, at 2.

97. 2016 PCAST REPORT, *supra* note 3, at 112.

98. *Id.* at 112 n.335. The meaning of 95 percent confidence is subtle and the description in the 2016 report is incorrect. David H. Kaye, *PCAST's Sampling Errors (Part I)*, FORENSIC SCI., STAT. & L. (Oct. 24, 2016), http://for-sci-law.blogspot.com/2016/10/pcasts-sampling-errors.html [https://perma.cc/L2L9-VHE9]. Another way to report the same estimate of a false declaration of a match when the materials tested did not come from the same gun is that this interval goes from the 0.96 percent to 2.30 percent. BALDWIN ET AL., *supra* note 84. For notes on some of the difficulties with PCAST's approach to estimating false-positive probabilities as measures of probative value in a particular case, see David H. Kaye, *PCAST's Sampling Errors (Part II: Getting More Technical)*, FORENSIC SCI., STAT. & L. (Dec. 11, 2016), http://for-sci-law.blogspot.com/2016/12/pcasts-sampling-errors-part-ii-getting.html [https://perma.cc/WX3H-XFD6]; David H. Kaye, *PCAST and the Ames Bullet Cartridge Study: Will the Real Error Rates Please Stand Up?*, FORENSIC SCI., STAT. & L. (Nov. 1, 2016), http://for-

But applying such numbers to individual examiners and particular cases is more challenging than the report recognizes. It is one thing to show that, as a group, some set of examiners can reach correct conclusions in comparisons that they do not regard as inconclusive. It is another to accurately estimate the probability of an error for a given examiner in a particular comparison.[99] Indeed, the 2016 report notes that "20 of the 22 false positives were made by just 5 of the 218 examiners—strongly suggesting that the false positive rate is highly heterogeneous across the examiners";[100] however, the report does not discuss the implications of this heterogeneity for testimony about "the error rates" that it wants "clearly presented."[101] It calls for "rigorous proficiency testing" of the examiner and disclosure of those test results.[102] There is a substantial argument for admitting both performance-test-based estimates of error rates, but the report does not

sci-law.blogspot.com/2016/11/pcast-and-ames-study-will-real-error.html [http s://perma.cc/T7JM-PY3Z].

99. This caveat does not mean that an average error rate in a study is irrelevant, or that only examiner-specific "proficiency tests" on casework-like samples of the same level of difficulty—in which examiner judgments also are analyzed as the output of a black-box system—are relevant. It is sensible to rely on average figures when nothing better is at hand, and to consider them in conjunction with an individual-specific error-rate even when one is available. *See generally* Dominique Fourdrinier & Martin T. Wells, *On Improved Loss Estimation for Shrinkage Estimators*, 27 STAT. SCI. 61, 61 (2012); Hermanus H. Lemmer, *Shrinkage Estimators*, *in* 12 ENCYCLOPEDIA OF STAT. SCI. 7704–7707 (Samuel Kotz, N. Balakrishnan, Campbell B. Read & Brani Vidakovic eds., 2d ed. 2006).

100. 2016 PCAST REPORT, *supra* note 3, at 110.

101. Baldwin et al. cautioned that:

> F]or the pool of participants used in this study the fraction of false positives was approximately 1%. The study was specifically designed to allow us to measure not simply a single number from a large number of comparisons, but also to provide statistical insight into the distribution and variability in false-positive error rates. The . . . overall fraction is not necessarily representative of a rate for each examiner in the pool. Instead, . . . the rate is a highly heterogeneous mixture of a few examiners with higher rates and most examiners with much lower error rates. This finding does not mean that 1% of the time each examiner will make a false-positive error. Nor does it mean that 1% of the time laboratories or agencies would report false positives, since this study did not include standard or existing quality assurance procedures, such as peer review or blind reanalysis.

BALDWIN ET AL., *supra* note 84, at 18.

102. 2016 PCAST REPORT, *supra* note 3, at 113.

develop the idea.[103] PCAST's discussion of a false-positive rate from a study designed to show whether examiners as a group are generally capable of reaching correct results without verification should not be taken as a final word on how to estimate error rates for courtroom use.[104]

VI. THE FUTURE

It seems unlikely that the PCAST report will result in the widespread judicial rejection of largely subjective comparisons.[105] But the recommendations and conclusions of yet a third body of accomplished scientists should intensify judicial reservations about testimony that the "chance of error [is] so remote as to be a 'practical impossibility.'"[106] If the report has this effect, the issue of how to present the evidence becomes more critical. As previously noted, phrases like "reasonable ballistic certainty" and "more likely than not" are not the

103. *See supra* note 99.

104. Verification by a second examiner also is relevant to presenting or using an error rate. As previously noted, if the errors occur independently across examiners (as might be the case if the verification is truly blind), then the relevant false-positive error rate from the Ames study drops to $(1.52\%)^2 = 0.0231\%$.

105. There are no published opinions on whether the analysis in the report warrants exclusion of firearm-mark evidence. In *United States v. Chester*, the district court wrote that "the report provides foundational scientific background and recommendations for further study. As such, the report does not dispute the accuracy or acceptance of firearm toolmark analysis within the courts." Order at 1–2, United States v. Chester, No. 13 CR 00774 (N.D. Ill. Oct. 7, 2016), ECF No. 875. This reasoning overlooks the report's insistence that the number of well-designed studies of firearm-mark identification is simply too small to establish the "foundational validity" required by Rule 702. Without explaining why PCAST's understanding of the extent of the research is wrong, the court added that the error rates in the Ames study and one of the other ones discussed in the report were "sufficiently low." *Id.* at 2.

106. 2016 PCAST REPORT, *supra* note 3, at 145 (recommending that courts should never permit scientifically indefensible claims such as: "'zero,' 'vanishingly small,' 'essentially zero,' 'negligible,' 'minimal,' or 'microscopic' error rates; '100 percent certainty' or proof 'to a reasonable degree of scientific certainty;' identification 'to the exclusion of all other sources;' or a chance of error so remote as to be a 'practical impossibility.'"). "Practical impossibility" and "practical certainty" are signature phrases for firearms examiners. *See supra* notes 12, 43 and accompanying text; *see also* ASS'N FIREARM & TOOLMARK EXAM'RS, *supra* note 75, at 1 ("[E]xaminers employing standard, validated procedures will rarely, *if ever*, commit false identifications or false eliminations.") (emphasis added).

solution.[107] Three more promising approaches are worth noting. If operating within the current paradigm of experience-and-training-based holistic conclusions, experts should not claim to be applying distinctly *scientific* methods for interpreting measurements or observations.[108] To follow the AFTE logic, they could explain that they have been trained in comparing the variations in the marks left by a gun, and that the marks seem to diverge from the normal range that they recall—but that they have no quantitative knowledge of the variation that normally exists when bullets are fired from the same gun as opposed to different guns.[109] And, any conclusion that the excess variation means that marks on the questioned item came from the known gun should be accompanied by meaningful error probabilities.

This kind of presentation corresponds to the "black box" perspective on the process. The examiner is treated no differently than a mysterious computer program that classifies questioned items into two categories—same gun, or different guns. The marks are the input or stimulus; a response of "same gun" or "different gun" is the output.[110] For the purpose of trusting the categorical conclusion, *how* the examiner performs the classification is not crucial.[111] The "operating characteristics" of the examiner as a source detector,[112] if adequately studied,

107. *See supra* notes 71–72 and accompanying text. Allowing testimony to "a reasonable degree of ballistic certainty," however, is a fig leaf that does not provide decent modesty. The witness often is presented as a scientist, applying a scientific method and using scientific terms. The phrase "to a reasonable degree of scientific certainty" adds nothing meaningful to the opinion of such a witness, and extirpating the phrase does not go far toward closing the distance between a firm opinion and a well-warranted one. KAYE ET AL., *supra* note 6, § 15.7.1 (Cum. Supp. 2015).

108. *See* Kaye, *supra* note 4, at 13–15.

109. As such, they should not use the phrases like "individual marks." *Cf.* KAYE ET AL., *supra* note 6, § 15.7.1(c), at 254 (Cum. Supp. 2015) ("The demand that the forensic science community perpetuate the time-honored but intellectually unsatisfying theory of individual versus class characteristics is unfortunate."). "Class characteristics" are acquired via a manufacturing or other process that is known to be uniform enough to produce many items with that characteristic. Other characteristics are acquired via a more variable process that produces fewer items with the same characteristic, but no law of nature dictates that an "individual characteristic" exists in one and only one item.

110. This is putting to the side a refusal to reach a clear conclusion by declaring that the evidence is inconclusive.

111. The fact that a classification procedure is based on a valid theory lends credence to its results, but the theory is not a complete substitute for empirical testing of the procedure or its components.

112. For a discussion of operating characteristics of a statistical classification procedure, see, for example, THOMAS D. WICKENS, ELEMENTARY SIGNAL DETECTION THEORY (2002); NAT'L RESEARCH COUNCIL, ASSEMBLY OF

are sufficient. Broadly speaking, this is the PCAST perspective on validation and presentation of traditional testimony.

However, it is not necessary for the examiner to be an inscrutable detector that registers either a same-gun signal or its absence. Many forensic scientists and statisticians favor a second mode of presentation in which the examiner describes (1) how often the perceived degree of agreement between the questioned specimen and those from the test firings would be seen if all the specimens came from the same gun and (2) how often such similarity would be seen if the questioned specimens came from a different gun.[113] The extent to which (1) exceeds (2) indicates how much the evidence supports the same-source conclusion as opposed to the different-source conclusion.[114] Describing the strength of the evidence in this manner—without any categorical conclusion from the expert's mind—is an attractive alternative to conventional testimony.[115] A firearms analyst should be able to articulate the "likelihoods"—the rough probabilities of the marks given each hypothesis about the source and the basis for these judgments about the evidence. Assessing the likelihoods is the expertise that lay jurors lack and that is supposed to come with training and experience in the field. But jurors can decide which likelihood ratios are large enough to warrant a source attribution as well as firearms experts can.[116] When experts take over that task, they end up presenting radically different conclusions for marks that are just shy of their implicit and

BEHAVIORAL AND SOC. SCIS., COMM. ON EVALUATION OF SOUND SPECTROGRAMS, ON THE THEORY AND PRACTICE OF VOICE IDENTIFICATION 27–30 (1979).

113. Geoffrey Stewart Morrison et al., *A Comment on the PCAST Report: Skip the "Match"/"Non-match" Stage*, 272 FORENSIC SCI. INT'L 7, 7–8 (2017); *see also* BERNARD ROBERTSON ET AL., INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM (2d ed. 2016); EUROPEAN NETWORK OF FORENSIC SCI. INSTS., ENFSI GUIDELINE FOR THE FORMULATION OF EVALUATIVE REPORTS IN FORENSIC SCIENCE: STRENGTHENING THE EVALUATION OF FORENSIC RESULTS ACROSS EUROPE (STEOFRAE) § 7.1.2–7.1.3, at 87–90 (2015); I. W. Evett et al., *Finding the Way Forward for Forensic Science in the US—A Commentary on the PCAST Report*, 278 FORENSIC SCI. INT'L 16 (2017).

114. *See e.g.*, KAYE ET AL., *supra* note 6, § 14.2.1, at 631; David H. Kaye, *Digging into the Foundations of Evidence Law*, 115 MICH. L. REV. 915, 923–25 (2017).

115. Of course, proof that examiners' judgments of the weight of evidence are reasonably accurate is necessary. *See, e.g.*, Kaye, *supra* note 4, at 21.

116. *Cf.* David H. Kaye, *Likelihoodism, Bayesianism, and a Pair of Shoes*, JURIMETRICS J., Fall 2012, at 1, 9 (applying this argument in the context of footwear-mark testimony).

unarticulated cutoff for source attribution than for marks that are barely over their threshold.[117]

The preceding two approaches are still predominantly subjective. In the longer term, we can and should expect expert testimony to be informed by statistical data about the frequency of types of marks on bullets or cartridge cases as determined from reference databases.[118] Three-dimensional imaging methods allow automated feature extraction.[119] With data on the distributions of similarity scores in items from the same gun and items from different ones, statistical models can generate quantitative likelihood ratios.[120] Such systems are statistically reliable—the same inputs generate the same outputs—and they can be validated empirically by investigating their performance on different data sets. Progress in these endeavors will enable firearms examiners to speak more fittingly of the "The Science Behind Firearm and Tool Mark Examination."[121]

117. *See e.g.*, ROBERTSON ET AL., *supra* note 113, § 5.4, at 63–64; Morrison et al., *supra* note 113, at 7–8.

118. In 2016, the National Institute of Standards and Technology established such a database. *NIST Ballistics Toolmark Database*, NAT'L INST. OF STANDARDS AND TECH. (Dec. 20, 2017), https://www.nist.gov/programs-projects/nist-ballistics-toolmark-database [https://perma.cc/3VGN-HGWT].

119. *See e.g.*, Daniel Ott et al., *Identifying Persistent and Characteristic Features in Firearm Tool Marks on Cartridge Cases*, SURFACE TOPOGRAPHY: METROLOGY & PROPS. (2017), http://iopscience.iop.org/article/10.1088/2051-@ben672X/aa864a [https://perma.cc/NEE7-MJGR].

120. *See e.g.*, Fabiano Riva & Christophe Champod, *Automatic Comparison and Evaluation of Impressions Left by a Firearm on Fired Cartridge Cases*, 59 J. FORENSIC SCI. 637, 638 (2014).

121. Nancy Ritter, *The Science Behind Firearm and Tool Mark Examination*, NAT'L INST. OF JUST. (Oct. 2014), https://nij.gov/journals/274/Pages/firearm-toolmark-examination.aspx [https://perma.cc/Z7BM-FQ5E].