THE HISTORY AND LEGACY OF THE LATENT FINGERPRINT BLACK BOX STUDY

BY **LUCAS ZARWELL** AND **GREGORY DUTTON**

The FBI's black box study on latent prints continues to influence the criminal justice system's understanding of the validity and reliability of forensic testimony.

A forensic scientist's testimony is vital for upholding justice in a court of law. The scientist's conclusions must be based on tested scientific methods with objective outcomes, without regard for whether the results may benefit the defense or the prosecution. Forensic methods are developed, measured, advanced, and evaluated through rigorous research — building a foundation for those conclusions to be evaluated and accepted by a court of law.

Examiner testimony — particularly in the forensic pattern disciplines (e.g., latent fingerprints, firearms, toolmarks, and footwear) — has been under heavy scrutiny in recent years. High-profile misidentifications, admissibility challenges, and blue-ribbon committee reports have heightened criticism about the scientific basis of examiner testimony in these disciplines and the forensic methods on which they are based.

"Black box" studies — those that measure the accuracy of outcomes absent information on how they are reached — can help the field better understand the validity and reliability of these methods. This article explores the basis of the black box design and highlights the history and legacy of one particularly influential study: a 2011 black box study by the FBI that examined the accuracy and reliability of latent fingerprint examiner decisions. This study had an immediate and lasting impact in the courts and continues to help define a path forward for future research. The article concludes with an overview of how the National Institute of Justice (NIJ) is working to support black box and similar studies across a number of forensic disciplines.

Misidentification in the Madrid Bombings

On March 11, 2004, attacks directly targeting commuter trains in Madrid, Spain, killed 193 people and injured approximately 2,000 others. On May 6, 2004, the FBI wrongfully arrested and detained Brandon Mayfield based on a latent fingerprint associated with the attacks. An official investigation later found that Mayfield, an American citizen from Washington County, Oregon, had no connection with the case. This led to a public apology from the FBI, internal reviews, and lawsuits to help compensate the wrongfully detained.

Sources: Matthew Harwood, "The Terrifying Surveillance Case of Brandon Mayfield," *Al Jazeera America,* February 8, 2014; and Office of the Inspector General, *A Review of the FBI's Handling of the Brandon Mayfield Case,* Executive Summary, Washington, DC: U.S. Department of Justice, Office of the Inspector General, Oversight and Review Division, January 2006, https://oig.justice.gov/sites/default/files/legacy/special/s0601/exec.pdf.

A Discipline Under Scrutiny

In 1993, the U.S. Supreme Court established five factors that a trial judge may consider when determining whether to admit scientific testimony in court.[1] Known as the *Daubert* standard, these factors are:

1. Whether the theory or technique in question can be and has been tested.

2. Whether it has been subjected to peer review and publication.

3. The degree of its known or potential error rate.

4. The existence and maintenance of standards controlling its operation.

5. Whether it has attracted widespread acceptance within a relevant scientific community.

One of the factors — a method's known or potential error rate — has arguably led to a substantial degree of confusion, discussion, and debate. This debate increased in 2004, when an appeals court in *United States v. Mitchell* recommended that, in future cases, prosecutors seek to show the individual error rates of expert witness examiners and not that of the forensic discipline in general.[2] The National Academy of Sciences has since addressed the confusion of practitioner error rates with discipline error rates; however, the scientific community continues to debate how to best define error rates overall.[3]

At about the same time as the *Mitchell* decision, an imbroglio resulting from an identification error involving the FBI's Latent Fingerprint Unit was unfolding. The misidentification — caused by an erroneous fingerprint individualization associated with the 2004 Madrid train bombings (see sidebar, "Misidentification in the Madrid Bombings") — led to a series of FBI corrective actions, including suspension of work, a two-year review of casework, and the establishment of an international review committee to evaluate the misidentification and make recommendations.[4]

In addition, the FBI Laboratory commissioned an internal review committee to evaluate the scientific basis of latent print examination and recommend research to improve our understanding of the discipline's validity. In 2006, the FBI committee found that the methodology surrounding latent fingerprint examination — like most pattern disciplines — has more subjectivity than other forensic disciplines, for example, chemical analysis of seized drugs. The FBI committee recommended black box testing, a technique to test both examiners and the methods used simultaneously.[5]

Black Box Testing

In his 1963 paper "A General Black Box Theory," physicist and philosopher Mario Bunge articulated a concept applied in software engineering, physics,

Exhibit 1. Black Box Testing

INPUT → Black Box → OUTPUT
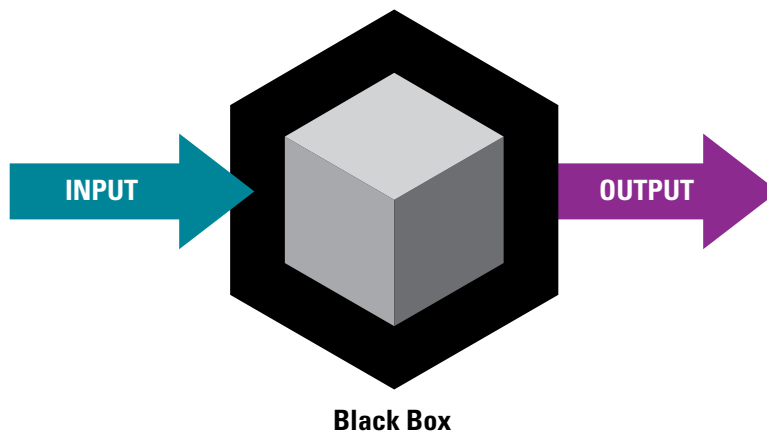
psychology, and other complex scientific systems.[6] Bunge represented a simplified black box as a notional system where inputs are entered and outputs emerge. Although the specific constitution and structure of the system are not considered, the system's overall behavior is accounted for.

Software validation offers one example of how a black box study can be applied. The tester may not know anything about the application's internal code; however, they have an expectation of a particular result based on the data provided. Another example is predicting consumer behavior. The consumer's thought processes are treated as a black box, and the study determines how they are likely to respond (i.e., will they purchase the item or not) when provided input from different marketing campaigns.

Today, this theory and its encompassing approach are being used to evaluate the reliability of forensic methods, measure their associated error rates, and give courts the information they need to assess the admissibility of the methods in question. A black box study measures the accuracy of examiners' conclusions without considering how they reached those conclusions. In essence, factors such as education, experience, technology, and procedure are all addressed as a single entity that creates a variable output based on input (see exhibit 1).

In 2011 — five years after the FBI committee's recommendation — Noblis (a scientific nonprofit)[7] and the FBI published the results of a black box study to examine the accuracy and reliability of forensic latent fingerprint decisions.[8] The discipline was found to be highly reliable and tilted toward avoiding false incriminations. The study reported a false positive rate of 0.1% and a false negative rate of 7.5%.[9] In other words, out of every 1,000 times examiners determined that two prints came from the same source, they were wrong only once. But when determining that two prints did *not* come from the same source, they were wrong nearly 8 out of 100 times. The report was introduced in court almost immediately after it was published,[10] and since then it has been well accepted by the scientific community. The report continues to be immensely influential; it has been downloaded more than 70,000 times and is among the top 5% of all research outputs in terms of impact online.[11] The research team went on to publish 15 additional papers delving deeper into aspects of latent print examination.[12]

In its 2016 report *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods,* the President's Council of Advisors on Science and Technology discussed the challenges in assessing the performance of both objective and subjective pattern comparison methods to determine if they are fit for purpose.[13] The council doubled down

Today, black box studies are being used to evaluate the reliability of forensic methods, measure their associated error rates, and give courts the information they need to assess the admissibility of the methods in question.

on the 2006 FBI research committee's conclusion by recommending similar black box studies for other forensic disciplines and cited the 2011 latent print study as an excellent example of how to accomplish this.

Why Was This Study So Effective?

There are several reasons why the FBI's latent print study was so successful. One key factor was the existing knowledge surrounding the science of latent fingerprint examination and its established historical application in the forensic sciences.

Latent print examination is a classic example of a forensic pattern discipline. In latent prints, the pattern being examined is formed by the fine lines that curve, circle, and arch on our fingertips, palms, and footpads. These lines are composed of grooves and friction ridges, which provide the traction that enables us to pick up a paperclip or quickly turn the page of a newspaper. However, they also leave impressions and residues that can be photographed or lifted from the surface of an item at a crime scene. These residues — formed by sweat, oils, and particulates — leave copies of the friction ridge patterns called "latents." Latent print examiners compare the ridge features of latent prints left at a crime scene to those collected under controlled conditions from a known individual. Controlled prints are called "exemplars" and are collected using ink on paper or a digital scanning device.

Today, the principal process used to examine latent prints is analysis, comparison, evaluation, and verification (ACE-V).[14] An examiner's subjective decisions are involved in the ACE component of the method, which involves:

1. Analyzing whether the quality of a latent print is good enough to be compared to an exemplar.

2. Comparing features of the latent print to the exemplar.

3. Evaluating the strength of that comparison.

The verification portion of the process involves a second examiner's independent analysis of the matched pair of prints.

ACE-V as typically implemented can yield four outcomes: no value (unsuitable for comparison), identification (originating from the same source), exclusion (originating from different sources), or inconclusive.[15] The verification step may be optional for exclusion or inconclusive decisions. For example, the Noblis/FBI latent print study applied the ACE portion of the process but did not include verification. This was a significant decision because excluding the verification step contributed to the upper bound for error rates reported by the study.[16] Nevertheless, the researchers were able to compare the conclusions of pairs of examiners to infer that verification likely could have prevented most errors.

There were a number of factors that made the study successful; other disciplines can and have adopted these factors. First, the FBI partnered with outside, independent researchers to design and perform the study. Noblis is a nonprofit science and technology organization with acumen in research and analysis. Together, the FBI and Noblis were a productive team — the FBI brought world-renowned expertise in latent print examination and forensic science research, and Noblis brought a reputation for objective analysis.

The relative size and scale of the study were also important. The FBI has a reputation for leadership and high-quality practices and training, and it actively contributes to practitioner professional groups

and meetings. The agency also had an extensive and transparent response to the 2004 Madrid misidentification, along with plans for future research. This reputation and approach helped broker trust from the forensic science community. As a result, more than 169 latent print examiners — from federal, state, and local agencies, as well as private practice — volunteered to be part of the study.[17] The scale of the study design was also large enough to produce statistically valid results. Each examiner compared approximately 100 print pairs out of a pool of 744 pairs, for a total of 17,121 individual decisions.[18]

In addition, the study was double-blind, open set, and randomized. Scientifically, these design elements are important because they mitigate potential bias. As a double-blind study, participants did not know the ground truth (the true match or nonmatch relationships) of the samples they received, and the researchers were unaware of the examiners' identities, organizational affiliations, and decisions. The open set of 100 fingerprint comparisons from a pool of 744 pairs[19] further strengthened the study by ensuring that not every print in an examiner's set had a corresponding mate. This prevented participants from using a process of elimination to determine matches. Finally, the randomized design varied the proportion of known matches and nonmatches across participants.

Lastly, the study design included a diverse range of quality and complexity. The study designers had latent print experts select pairs from a much larger pool of images that included broad ranges of print quality and comparison difficulty.[20] They intentionally included challenging comparisons, so that the error rates measured would represent an upper limit for the errors encountered in real casework.

Impact on the Courts

The major impact of black box research has been in the courts. Following publication, the results of the FBI latent print black box study were almost immediately applied in an opinion to deny a motion to exclude FBI latent print evidence.[21] The case involved a bombing

at the Edward J. Schwartz federal courthouse in San Diego. Donny Love, Sr., with the help of his accomplices, masterminded the construction and placement of several explosive devices, one of which was used to bomb the federal courthouse. Although no one was injured or killed, the explosion blew out the doors to the federal courthouse and sent shrapnel and nails flying over a block away and at least six stories into the air.[22]

In the motion, Love argued that latent fingerprint analysis was insufficiently reliable for admission under Federal Rule of Evidence 702 and the Supreme Court's previous opinions in *Daubert v. Merrell Dow Pharmaceuticals* (1993) and *Kumho Tire Company v. Carmichael* (1999). Therefore, Love argued, the analyst's testimony about the latent prints she analyzed for this case was also insufficiently reliable for admission.

The FBI latent print study results were entered into the record supporting latent print examination and cited explicitly in the opinion when considering the method's reliability under factor 3 of the *Daubert* standard (known or potential error rates). In the opinion, which led to the denial of the motion to exclude and an eventual guilty verdict, the judge stated, "All of the relevant evidence in the record before the court suggests that the ACE-V methodology results in very few false positives — which is to say, very few cases in which an examiner identifies a latent print as matching a known print even though different individuals made the two prints."[23] The judge continued, "Most significantly, the May 2011 study of the performance of 169 fingerprint examiners revealed a total of six false positives among 4,083 comparisons of non-matching fingerprints for 'an overall false-positive rate of 0.1%.'"[24]

Other important rulings followed. *United States v. McCluskey* (2013) involved the double murder of Gary and Linda Hass, who had been shot and burned inside their travel trailer in August 2010.[25] The individuals charged with the crime — now both convicted — had left their fingerprints on a piece of plastic wrapper inside a pickup truck they stole from the murdered couple.[26] At trial, the defense issued a motion to

exclude fingerprint evidence and requested a *Daubert* hearing. One basis for the defense argument was the 2009 National Research Council report that stated, "There is no systematic, controlled validation study that purports to estimate the accuracy of latent print individualization."[27] In response, the court's opinion cited the FBI latent print study extensively to demonstrate *Daubert* factor 1 (the theory can be tested) and factor 3 (known or potential error rates). The opinion stated, "While the Brandon Mayfield case, along with other weaknesses in fingerprint testing, may provide fertile ground for cross-examination of the Government's fingerprint identification expert, it alone does not outweigh the testing that has been conducted in this area."[28]

Three years later, in *United States v. Fell* (2016),[29] an individual who was sentenced to death in 2006 for carjacking and death resulting from kidnapping and carjacking was seeking dismissal of the prior conviction based on the unreliability of fingerprint evidence. His fingerprints had been found in the car used in the kidnapping. The judicial opinion on the *Daubert* challenge to admit the fingerprint evidence cited the error rates determined in the FBI's latent print study, as well as subsequent research supporting examiner accuracy. This included studies exploring the repeatability and reproducibility of examiner conclusions and measuring how much information an examiner needs to make an identification.[30]

The Study's Legacy

The FBI's latent print black box study — with its robust design and transparent results — has spawned additional research in latent prints that explores the reproducibility and repeatability of examiner decisions, assesses quality and clarifying information, and explores interexaminer decisions.[31]

This landmark study has also influenced research in other forensic pattern disciplines, including palm prints, bloodstain patterns, firearms, handwriting, footwear, and, most recently, tire tread and digital evidence.[32] Black box studies in these disciplines present different challenges from latent prints. For

example, firearms examiners face a variety of makes and models of firearms that mark casings and bullets differently. This leads to diverse class and subclass characteristics in addition to individualizing features.[33] Within some disciplines, such as bloodstain pattern analysis, a range of practices and terminology currently exist; community consensus and uniform standards may be needed.

Even with these challenges, court decisions demonstrate the continued importance of black box studies. For example, in a motion to exclude ballistic evidence from a felony firearm possession case, the court in *United States v. Shipp* (2019)[34] cited a 2014 firearms black box study.[35] The court relied on the study's assessment that it most closely followed conditions that might be encountered in casework. The court noted, however, that the study demonstrated that a firearms toolmark examiner may "incorrectly conclude that a recovered piece of ballistics evidence matches a test fire once out of every 46 examinations" and "when compared to the error rates of other branches of forensic science — as rare as 1 in 10 billion for single source or simple mixture DNA comparisons … — this error rate cautions against the reliability of the [method]."[36] As a result, the court did not exclude the evidence but rather concluded that the examiner "will be permitted to testify only that the toolmarks on the recovered bullet fragment and shell casing are consistent with having been fired from the recovered firearm."[37] Thus, the recovered firearm could not be excluded as a source, but the examiner would not be allowed to specifically associate the evidence to that individual firearm.

Black box studies of examiner conclusions have been and will continue to be important to our understanding of the validity and reliability of forensic testimony, especially in the pattern comparison disciplines. Further studies — modeled on the FBI latent print study design and involving relevant practitioner communities — will provide value to courts considering *Daubert* challenges to admissibility. NIJ continues to support black box and similar studies across a number of forensic disciplines.[38] Explore the projects below for more information:

- "A Black Box Study of the Accuracy and Reproducibility of Tire Evidence Examiners' Conclusions," award number 2020-DQ-BX-0026.

- "Inter-Laboratory Variation in Interpretation of DNA Mixtures," award number 2020-R2-CX-0049.

- "Black Box and White Box Forensic Examiner Evaluations — Understanding the Details," award number DJO-NIJ-19-RO-0010.

- "Black Box Evaluation of Bloodstain Pattern Analysis Conclusions," award number 2018-DU-BX-0214.

- "Firearm Forensics Black-Box Studies for Examiners and Algorithms Using Measured 3D Surface Topographies," award number 2017-IJ-CX-0024.

- "Testing the Accuracy and Reliability of Palmar Friction Ridge Comparisons: A Black Box Study," award number 2017-DN-BX-0170.

- "Kinematic Validation of FDE Determinations About Writership in Questioned Handprinting and Handwriting," award number 2017-DN-BX-0148.

- "Understanding the Expert Decision-Making Process in Forensic Footwear Examinations: Accuracy, Decision Rules, Predictive Value, and the Conditional Probability of an Outcome," award number 2016-DN-BX-0152.

About the Authors

Lucas Zarwell, M.F.S., is the director of NIJ's Office of Investigative and Forensic Sciences. **Gregory Dutton,** Ph.D., is a program manager in NIJ's Office of Investigative and Forensic Sciences.

Notes

1. *Daubert v. Merrell Dow Pharm., Inc.,* 509 U.S. 579, 113 S. Ct. 2786 (1993).

2. *United States v. Mitchell,* 365 F.3d 215 (3d Cir. 2004).

3. National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward,* Washington, DC: The National Academies Press, 2009, https://doi.org/10.17226/12589.

4. Robert B. Stacey, "Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case," *Forensic Science Communications* 7 no. 1 (2005), https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2005/special_report/2005_special_report.htm; Office of the Inspector General, *A Review of the FBI's Handling of the Brandon Mayfield Case,* Washington, DC: U.S. Department of Justice, Office of the Inspector General, Oversight and Review Division, March 2006, https://oig.justice.gov/sites/default/files/archive/special/s0601/final.pdf; and Office of the Inspector General, *A Review of the FBI's Progress in Responding to the Recommendations in the Office of the Inspector General Report on the Fingerprint Misidentification in the Brandon Mayfield Case,* Washington, DC: U.S. Department of Justice, Office of the Inspector General, Oversight and Review Division, June 2011, https://www.oversight.gov/sites/default/files/oig-reports/s1105.pdf.

5. Bruce Budowle, JoAnn Buscaglia, and Rebecca Schwartz Perlman, "Review of the Scientific Basis for Friction Ridge Comparisons as a Means of Identification: Committee Findings and Recommendations," *Forensic Science Communications* 8 no. 1 (2006), https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2006/research.

6. Mario Bunge, "A General Black Box Theory," *Philosophy of Science* 30 no. 4 (1963): 346-358, https://doi.org/10.1086/287954.

7. Noblis is an independent, nonprofit organization that serves government through scientific and technical expertise. Noblis has deep research experience in modeling, simulation, data analytics, and life sciences — all of which contributed to designing a successful research design. For more information, see https://noblis.org/what-we-do/.

8. Bradford T. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions," *Proceedings of the National Academy of Sciences of the United States of America* 108 no. 19 (2011): 7733-7738, https://doi.org/10.1073/pnas.1018707108.

9. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

10. *United States v. Love,* No. 10cr2418-MMM, 2011 U.S. Dist. LEXIS 53213 (S.D. Cal. May 17, 2011).

11. According to Altmetrics, https://pnas.altmetric.com/details/102412023.

12. Thomas A. Busey et al., "Characterizing Missed Identifications and Errors in Latent Fingerprint Comparisons Using Eye-Tracking Data," *PLoS ONE* 16 no. 5 (2021): e0251674, https://doi.org/10.1371/journal.pone.0251674; Nathan D. Kalka, Michael Beachler, and R. Austin Hicklin, "LQMetric: A Latent Fingerprint Quality Metric for Predicting AFIS Performance and Assessing the Value of Latent

Fingerprints," *Journal of Forensic Identification* 70 no. 4 (2020): 443-463; R. Austin Hicklin et al., "Why Do Latent Fingerprint Examiners Differ in Their Conclusions?" *Forensic Science International* 316 (2020): 110542, https://doi.org/10.1016/j.forsciint.2020.110542; R. Austin Hicklin et al., "Gaze Behavior and Cognitive States During Fingerprint Target Group Localization," *Cognitive Research: Principles and Implications* 4 no. 12 (2019), https://doi.org/10.1186/s41235-019-0160-9; Bradford T. Ulery et al., "Factors Associated With Latent Fingerprint Exclusion Determinations," *Forensic Science International* 275 (2017): 65-75, https://doi.org/10.1016/j.forsciint.2017.02.011; Bradford T. Ulery et al., "Data on the Interexaminer Variation of Minutia Markup on Latent Fingerprints," *Data in Brief* 8 (2016): 158-190, https://doi.org/10.1016/j.dib.2016.04.068; Bradford T. Ulery et al., "Interexaminer Variation of Minutia Markup on Latent Fingerprints," *Forensic Science International* 264 (2016): 89-99, http://dx.doi.org/10.1016/j.forsciint.2016.03.014; Bradford T. Ulery et al., "Changes in Latent Fingerprint Examiners' Markup Between Analysis and Comparison," *Forensic Science International* 247 (2015): 54-61, https://doi.org/10.1016/j.forsciint.2014.11.021; R. Austin Hicklin et al., "In Response to Haber and Haber, 'Experimental Results of Fingerprint Comparison Validity and Reliability: A Review and Critical Analysis,'" *Science and Justice* 54 no. 5 (2014): 390-391, https://doi.org/10.1016/j.scijus.2014.06.007; Bradford T. Ulery et al., "Measuring What Latent Fingerprint Examiners Consider Sufficient Information for Individualization Determinations," *PLoS ONE* 9 no. 11 (2014): e110179, https://doi.org/10.1371/journal.pone.0110179; Nathan D. Kalka and R. Austin Hicklin, "On Relative Distortion in Fingerprint Comparison," *Forensic Science International* 244 (2014): 78-84, https://doi.org/10.1016/j.forsciint.2014.08.007; Bradford T. Ulery et al., "Understanding the Sufficiency of Information for Latent Fingerprint Value Determinations," *Forensic Science International* 230 nos. 1-3 (2013): 99-106, https://doi.org/10.1016/j.forsciint.2013.01.012; R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts, "Assessing the Clarity of Friction Ridge Impressions," *Forensic Science International* 226 nos. 1-3 (2013): 106-117, https://doi.org/10.1016/j.forsciint.2012.12.015; Bradford T. Ulery et al., "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners," *PLoS ONE* 7 no. 3 (2012), https://doi.org/10.1371/journal.pone.0032800; R. Austin Hicklin et al., "Latent Fingerprint Quality: A Survey of Examiners," *Journal of Forensic Identification* 61 no. 4 (2011): 385-419; and Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

13. President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods,* Washington, DC: Executive Office of the President, President's Council of Advisors on Science and Technology, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

14. Organization of Scientific Area Committees (OSAC) for Forensic Science, "OSAC Standard Framework for Developing Discipline Specific Methodology for ACE-V," 2020, https://www.nist.gov/system/files/documents/2020/04/06/Virtual%20Sub%203%20ACE-V%20Process%20Map_CDi_JAN2020.pdf.

15. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

16. In other words, if the verification step had been included in the experiment, its presence may have rightfully eliminated any detectable error in the study. This is desirable when a process is intended to confirm the identity of a suspect; however, it would disintegrate any error the study was designed to detect.

17. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

18. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

19. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

20. Ulery et al., "Accuracy and Reliability of Forensic Latent Fingerprint Decisions."

21. *United States v. Love*.

22. U.S. Attorney's Office, Southern District of California, "Final Defendant in San Diego Federal Courthouse Bombing Sentenced," press release, San Diego: U.S. Attorney's Office, Southern District of California, February 15, 2013, https://archives.fbi.gov/archives/sandiego/press-releases/2013/final-defendant-in-san-diego-federal-courthouse-bombing-sentenced.

23. *United States v. Love.*

24. *United States v. Love.*

25. Federal Bureau of Investigation, "Violent Criminals Sentenced: Charged in 2010 Murder of Oklahoma Couple," July 3, 2014, https://www.fbi.gov/news/stories/-sentence-in-2010-murder-of-oklahoma-couple.

26. *United States v. McCluskey,* No. 10-2734 JCH, 2013 U.S. Dist. LEXIS 202828 (D.N.M. July 22, 2013).

27. National Research Council of the National Academies, *Strengthening Forensic Science in the United States.*

28. *United States v. McCluskey.*

29. *United States v. Fell,* No. 5:01-cr-12-01, 2016 U.S. Dist. LEXIS 198728 (D. Vt. Sep. 13, 2016).

30. Ulery et al., "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners"; and Ulery et al., "Measuring What Latent Fingerprint Examiners Consider Sufficient Information for Individualization Determinations."

31. For example, Hicklin, Buscaglia, and Roberts, "Assessing the Clarity of Friction Ridge Impressions"; and Ulery et al., "Interexaminer Variation of Minutia Markup on Latent Fingerprints."

32. Heidi Eldridge, Marco De Donno, and Christophe Champod, "Testing the Accuracy and Reliability of Palmar Friction Ridge Comparisons – A Black Box Study," *Forensic Science International* 318 (2021): 110457, https://doi.org/10.1016/j.forsciint.2020.110457; National Institute of Justice funding award description, "Black Box Evaluation of Bloodstain Pattern Analysis Conclusions," at Noblis, Inc., award number 2018-DU-BX-0214, https://nij.ojp.gov/funding/awards/2018-du-bx-0214; Thomas G. Fadul Jr. et al, "An Empirical Study To Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing Consecutively Manufactured Glock EBIS Barrels With the Same EBIS Pattern," Final report to the National Institute of Justice, award number 2010-DN-BX-K269, December 2013, NCJ 244232, https://nij.ojp.gov/library/publications/empirical-study-improve-scientific-foundation-forensic-firearm-and-tool-mark-0; Jacqueline A. Speir, Nicole Richetelli, and Lesley Hammer, "Forensic Footwear Reliability: Part I — Participant Demographics and Examiner Agreement," *Journal of Forensic Sciences* 65 no. 6 (2020): 1852-1870, https://doi.org/10.1111/1556-4029.14553; Nicole Richetelli, Lesley Hammer, and Jacqueline A. Speir, "Forensic Footwear Reliability: Part II — Range of Conclusions, Accuracy, and Consensus," *Journal of Forensic Sciences* 65 no. 6 (2020): 1871-1882, https://doi.org/10.1111/1556-4029.14551; Nicole Richetelli, Lesley Hammer, and Jacqueline A. Speir, "Forensic Footwear Reliability: Part III — Positive Predictive Value, Error Rates, and Inter-Rater Reliability," *Journal of Forensic Sciences* 65 no. 6 (2020): 1883-1893, https://doi.org/10.1111/1556-4029.14552; National Institute of Justice funding award description, "A Black Box Study of the Accuracy and Reproducibility of Tire Evidence Examiners' Conclusions," at Noblis, Inc., award number 2020-DQ-BX-0026, https://nij.ojp.gov/funding/awards/2020-dq-bx-0026; Chad Chapnick et al., "Results of the 3D Virtual Comparison Microscopy Error Rate (VCMER) Study for Firearm Forensics," *Journal of Forensic Sciences* 66 no. 2 (2021): 557-570, https://doi.org/10.1111/1556-4029.14602; and National Institute of Standards and Technology, "NIST to Digital Forensics Experts: Show Us What You Got," June 2, 2020, https://www.nist.gov/news-events/news/2020/06/nist-digital-forensics-experts-show-us-what-you-got. See also Brian C. McVicker et al., "A Method for Characterizing Questioned Footwear Impression Quality," *Journal of Forensic Identification* 71 no. 3 (2021): 205-216; L. Scott Chumbley et al., "Accuracy, Repeatability, and Reproducibility of Firearm Comparisons, Part 1: Accuracy," https://arxiv.org/ftp/arxiv/papers/2108/2108.04030.pdf; and Keith L. Monson, Erich D. Smith, and Stanley J. Bajic, "Planning, Design, and Logistics of a Decision Analysis Study: The FBI/Ames Study Involving Forensic Firearms Examiners," *Forensic Science International: Synergy* 4 (2022): 100221, https://doi.org/10.1016/j.fsisyn.2022.100221.

33. A class characteristic is "a feature shared by two or more items of footwear or tires. The footwear outsole or tire tread design and the physical size features of a footwear outsole or tire tread are two common manufactured class characteristics. General wear of the outsole or tire tread is also a class characteristic." OSAC Lexicon, "Class Characteristic," Organization of Scientific Area Committees for Forensic Science, https://lexicon.forensicosac.org.

34. *United States v. Shipp,* 422 F. Supp. 3d 762, 2019 U.S. Dist. LEXIS 205397.

35. David P. Baldwin et al., "A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons," U.S. Department of Energy, Ames Laboratory, Technical Report #IS-5207, 2014, https://www.ojp.gov/pdffiles1/nij/249874.pdf.

36. *United States v. Shipp.*

37. *United States v. Shipp.*

38. NIJ contributed to the FBI/Ames firearms study and the FBI/Noblis studies on handwriting and shoeprints.

Image source: pkripper503/iStock.com.