

Forensic Science and the Judicial Conformity Problem

*Katie Kronick**

I. INTRODUCTION	590
II. HOW <i>DAUBERT</i> AND <i>FRYE</i> HAVE FAILED TO KEEP OUT FAULTY FORENSIC SCIENCE DESPITE EVIDENCE OF FLAWS	596
A. Reports on the Reliability of Forensic Sciences Demonstrate Substantial Problems.....	597
1. Ballistic Imaging Report	597
2. NAS Report.....	598
3. PCAST Report and the National Commission on Forensic Science	600
B. Daubert, Frye, and Forensic Science	602
1. <i>Frye</i>	602
2. <i>Daubert</i>	603
C. The Limited Reliability of Firearm and Toolmark Examination	605
III. JUDICIAL DECISION-MAKING, FORENSIC SCIENCE, AND THE CONFORMITY PROBLEM.....	609
A. The Impact of Internal Cognitive Biases on Judicial Decision-Making.....	610
B. The Relationship Among External Conformity Pressures, Judicial Decision-Making, and Forensic Science.....	613
1. The Conformity Problem in the Context of Judicial Decision-Making.....	615
2. Forensic Science as a Strongly Held Belief.....	621
3. The Role of Prosecutors in Maintaining the Conformity Problem.....	623
IV. INSTANCES OF THE CONFORMITY PROBLEM IN JUDICIAL DECISION-MAKING	629
A. New Jersey Identification Cases.....	630

*Practitioner-in-Residence at American University's Washington College of Law. Many thanks to Jenny Roberts, Andrew Ferguson, Vida Johnson, Jessica Millward, Maneka Sinha, Valena Beety, and Michael Wilson for their assistance in developing this project. I am also grateful for feedback provided at the 2019 *Clinical Law Review* Writers' Workshop. Special thanks to Sarah Escobar for tireless research help.

B. District of Columbia Cases.....	636
C. District of Massachusetts Cases	638
V. CONCLUSION.....	642

I. INTRODUCTION

In a criminal case, forensic science can answer the unanswered questions and tie up the loose ends. DNA testing can tell us whose sweat is found in the gloves left at a crime scene. Chemical analysis can tell us whether the powder found in a small bag is cocaine or baking soda. We use fingerprint analysis to identify who has been in a stolen car and firearm and toolmark examination to identify whether a bullet casing at the scene of a homicide came from a specific gun. As powerful as this evidence can be, unfortunately, it can be riddled with errors resulting in faulty or even completely false information going before a jury. The errors can be unintentional human error, such as contamination at the testing stage or cognitive biases.¹ They can be intentional human error, such as claims that items were tested and produced a certain result when no testing ever occurred.² Or the errors can be foundational to the

¹ See, e.g., Itiel E. Dror, *A Hierarchy of Expert Performance*, 5 J. APPLIED RES. MEMORY & COGNITION 121, 122–23 (2016); Itiel E. Dror et al., *When Emotions Get the Better of Us: The Effect of Contextual Top-Down Processing on Matching Fingerprints*, 19 APPLIED COGNITIVE PSYCHOL. 799, 806–07 (2005).

² See, e.g., Keith L. Alexander, *Federal Prosecutors Question ‘Integrity and Competence’ of D.C. Crime Lab*, WASH. POST (Feb. 11, 2020), https://www.washingtonpost.com/local/public-safety/federal-prosecutors-question-integrity-and-competence-of-dc-crime-lab/2020/02/11/9135af14-48d2-11ea-bdbf-1dfb23249293_story.html?fbclid=IwAR3Dffztdp1CoMUTOc9dIwDGRCrUR6lrktMGVdE96tNYekiervwYtbK1h5g (detailing a recent letter that the U.S. Attorney’s Office for the District of Columbia sent to city officials requesting further investigation into issues with the crime lab, including problems with the leadership of the lab); Nick Bunkley, *Detroit Police Lab Is Closed After Audit Finds Serious Errors in Many Cases*, N.Y. TIMES (Sept. 25, 2008), <https://www.nytimes.com/2008/09/26/us/26detroit.html> (discussing the closure of a Detroit police laboratory after an audit found that, among other issues, the laboratory was only in compliance with 42% of the “essential standards” when compliance with 100% is required); S.P. Sullivan, *State Police Lab Scandal Led to Major Overhaul in How N.J. Tests Drug Evidence. It Was All Based on a Lie, Lawsuit Claims*, NJ.COM (Feb. 4, 2019), <https://www.nj.com/politics/2019/02/state-police-lab-scandal-led-to-major-overhaul-in-how-nj-tests-drug-evidence-it-was-all-based-on-a-lie-lawsuit-claims.html> (discussing the falsification of drug test results by a New Jersey lab technician that led to a change in drug testing procedures in the state). A recent review of the exonerations from 2018 revealed that a prominent cause of wrongful convictions was false or misleading testimony from forensic science experts, particularly where they exaggerated the degree of certainty to which they could testify. See *Exonerations in 2018*, NAT’L REGISTRY OF EXONERATIONS (Apr. 9, 2019), <https://www.law.umich.edu/special/exoneration/Documents/Exonerations%20in%202018.pdf> (discussing exonerations across the country in 2018 and patterns in the causes of wrongful convictions); Heather Murphy, *A Leading Cause for Wrongful Convictions: Experts Overstating Forensic Results*,

forensic science, such as bullet lead examination, which was once thought to be able to match a single bullet to a set of bullets but has been proven to lack evidentiary value.³

Regardless of the cause of the error, faulty forensic science has devastating effects. Of the 2,705 exonerations since 1989,⁴ almost a quarter of those cases involved faulty forensic science evidence.⁵ These numbers, however, do not account for the many wrongful convictions that will never be uncovered.⁶ The bottom line is that error is error. People are harmed whether the error is due to a lab's gross negligence, a well-meaning forensic analyst overstating results, or a fundamentally flawed understanding of the science.

These exonerations and the studies and analyses that followed have revealed some areas of forensic science as being particularly unreliable, including bitemark analysis and hair microscopy.⁷ Yet in the

N.Y. TIMES (Apr. 20, 2019), <https://www.nytimes.com/2019/04/20/us/wrongful-convictions-forensic-results.html> (discussing the issues with misleading forensic science in the trials of those wrongfully convicted and exonerated in 2018).

³ In 2004, the National Research Council released a report demonstrating that bullet lead examination, also known as Compositional Analysis of Bullet Lead ("CABL"), is not sufficiently reliable as standalone evidence. See NAT'L RES. COUNCIL, FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE (2004), <https://www.nap.edu/read/10924/chapter/2#4>. Essentially, this report found that though bullets with a similar composition may have come from the same batch, these batches can be as large as 35 million, making the evidentiary value almost nonexistent. See *id.* at 6–7.

⁴ *Exonerations in the United States Map*, NAT'L REGISTRY OF EXONERATIONS, <http://www.law.umich.edu/special/exoneration/Pages/Exonerations-in-the-United-States-Map.aspx> (last visited Dec. 15, 2020). This number is current as of November 15, 2020, but it will continue to increase as more innocent people are exonerated.

⁵ *Id.*

⁶ For example, a recent study estimates that 4.1% of death row inmates would be exonerated if they remained on death row indefinitely—only 2.3% have, in fact, been exonerated (based on data from 1973–2004). Samuel R. Gross et al., *Rate of False Conviction of Criminal Defendants Who Are Sentenced to Death*, 111 PROC. NAT'L ACAD. SCI. U.S. 7230, 7230–31, 7234 (2014). The possible rate of wrongful convictions outside the death penalty context, however, is impossible to know because far less litigation occurs post-conviction in non-capital cases, and those cases often do not involve DNA evidence, one of the primary tools of identifying wrongful convictions. "False convictions, by definition, are unobserved when they occur: If we know that a defendant is innocent, he is not convicted in the first place.... As a result, the great majority of innocent defendants remain undetected." *Id.* at 7230; see also Jenny Roberts, *The Innocence Movement and Misdemeanors*, 98 B.U.L. REV. 779, 784 (2018) (exploring the new interest in wrongful convictions at the misdemeanor level and noting that "[w]ithout a doubt, the lower criminal courts convict many innocent people of misdemeanors").

⁷ See, e.g., COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT'L RSCH. COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 42 (2009) [hereinafter NAS REPORT]; PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFF. OF THE PRESIDENT, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016), [hereinafter PCAST REPORT], <https://obamawhite>

not too distant past, “experts” in these areas were testifying that they could match a strand of hair or a bitemark to a particular individual to the exclusion of all others,⁸ an assertion we now know to be false. Today, some of the very experts who previously testified to the infallibility of these “sciences” are the ones asserting they should not be used.⁹

A 2009 report from the National Academy of Sciences (the “NAS Report”) found that before the use of DNA in criminal matters, which started in 1986, the only evaluation of whether forensic science evidence was reliable was cross-examination in the courtroom.¹⁰ “[A]lthough the precise error rates of these forensic tests are still unknown, comparison of their results with DNA testing in the same cases has revealed that some of these analyses, as currently performed, produce erroneous results.”¹¹ The resulting consequence is that innocent people are convicted of crimes they did not commit.¹²

Even though hundreds of wrongful convictions have been based on faulty forensic science evidence,¹³ once-trusted forensic science is now known to have been based on mere superstition,¹⁴ and the scientific community is calling for caution when relying on this evidence,¹⁵ judges routinely allow forensic science evidence to go before juries, often in

house.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

⁸ This is called “individualization”: the ability to take an item and match it to a single source to the exclusion of all other possible sources, such as claiming that a fingerprint on a gun matches the suspect in a case. Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 Sci. 892, 892 (2005). Fingerprint analysis, firearm and toolmark examination, bitemark analysis, hair microscopy, and DNA all purport to individualize (though DNA evidence is the only one to have the scientific evidence substantiating the claims). *Id.*

⁹ See, e.g., Liliana Segura & Jordan Smith, *Forensic Experts Fight Over the Problem of Junk Science*, INTERCEPT (May 5, 2019, 8:00 AM), <https://theintercept.com/2019/05/05/forensic-evidence-aafs-junk-science> (describing how Dr. Frank Wright, a forensic odontologist, who previously touted the use of bite-mark analysis, has come to the conclusion that he was wrong).

¹⁰ See NAS REPORT, *supra* note 7 (discussing the lack of scientific scrutiny placed on forensic sciences developed in crime laboratories).

¹¹ *Id.*

¹² See *id.* at 42–43.

¹³ See e.g., *Forensic Science: Problems and Solutions*, INNOCENCE PROJECT, <https://innocenceproject.org/forensic-science-problems-and-solutions> (last visited Nov. 13, 2020).

¹⁴ David S. Caudill, *Toward a Sociology of Forensic Knowledge? A (Supplementary) Response to Cole*, 48 SETON HALL L. REV. 583, 599 (2018); Mark Hansen, *Long-Held Beliefs About Arson Science Have Been Debunked After Decades of Misuse*, ABA J. (Dec. 1, 2015), https://www.abajournal.com/magazine/article/long_held_beliefs_about_arson_science_have_been_debunked_after_decades_of_m.

¹⁵ See e.g., NAS REPORT, *supra* note 7.

cases where defendants face very serious charges.¹⁶ Not only are judges willing to admit this evidence but they will frequently do so after denying a defense request for a hearing on the admissibility of such evidence under either the *Frye*¹⁷ or *Daubert*¹⁸ standards and, more often than not, without any critical analysis.¹⁹ Judges are supposed to provide a gatekeeping function, but when it comes to forensic science in criminal cases, the gates seem to be perennially open.²⁰

One might assume that judges would be terrified of presiding over a trial that results in a wrongful conviction, particularly in a case where the stakes are high, such as homicide or sexual assault. For example, Robert Lee Stinson was wrongfully convicted of sexual assault and murder based on faulty bitemark analysis. Years later, the judge who presided over the trial, former Wisconsin Supreme Court Justice Geske, stated, “I think I did everything right at the time. But that’s not to say that I don’t feel terrible.”²¹ She went on to explain that, at the time, the evidence was helpful to the jury, nothing said the evidence was not reliable, and it was relevant to the issues.²² Now though, after Mr.

¹⁶ Jane Campbell Moriarty, *Deceptively Simple: Framing, Intuition, and Judicial Gatekeeping of Forensic Feature-Comparison Methods Evidence*, 86 FORDHAM L. REV. 1687, 1694–95 (2018) (“Despite robust challenges to [feature-comparison methods] evidence over the last three decades, courts have rebuffed nearly all claims”) (internal citations omitted); Jennifer Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 BROOK. L. REV. 1209, 1213 (2010).

¹⁷ *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (establishing general acceptance as the standard for admitting novel scientific evidence). For further discussion, see *infra* Section II.B.1.

¹⁸ *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 587–95 (1993) (finding that Rule 702 of the Federal Rules of Evidence superseded the *Frye* standard of general acceptance in federal courts and establishing guidelines for trial courts to use in determining whether the scientific evidence in question is scientifically valid and can be properly applied to the facts of the case). For further discussion, see *infra* Section II.B.2.

¹⁹ Though the federal system and some states use the *Daubert* standard for determining whether expert testimony is admissible, many states continue to use the *Frye* standard. For a general discussion of how these standards are applied, see *infra* Section II.B.

²⁰ See Deidre Dwyer, *(Why) Are Civil and Criminal Expert Evidence Different?*, 43 TULSA L. REV. 381, 383 (2007) (noting that “the expert evidence of criminal prosecutors is subject to less scrutiny than that of criminal defendants, or than that of civil parties”). Several studies examining how judges evaluate expert testimony demonstrate that judges are more likely to admit evidence the prosecutor puts forward despite defense challenges, and where the defense seeks to admit expert testimony, the court is more likely to exclude such evidence. See Stephanie L. Damon-Moore, Comment, *Trial Judges and the Forensic Science Problem*, 92 N.Y.U. L. REV. 1532, 1557–58 (2017) (discussing post-*Daubert* rates of success in challenging expert evidence).

²¹ Joss Fong & Dion Lee, *False Positive*, Vox (Feb. 14, 2019, 12:00 PM), <https://www.vox.com/videos/2019/2/14/18223871/wrongful-conviction-bite-marks-robert-lee-stinson>.

²² *Id.*

Stinson served twenty-three years for a crime he did not commit, she “felt bad . . . [and was] just grateful we don’t have a death penalty [in Wisconsin].”²³ Justice Geske’s regret in response to presiding over a trial that resulted in a wrongful conviction appears genuine. In decision after decision regarding forensic science evidence, however, judges do not appear concerned about the presentation of faulty forensic science evidence or that its presentation could result in yet another wrongful conviction.

How is it that judges continue to allow these cases to go forward, to allow for the presentation of possibly faulty evidence, to risk participating in the conviction of a person who had nothing to do with the offense? Quite simply, a courtroom might be the perfect storm of influences that compel a judge to follow the norm (admit the evidence) rather than appropriately deviate (prohibit the unreliable evidence or limit its admissibility).

This Article explores whether pressure on judges to conform with their peers is a contributing influence on their decision-making concerning forensic science. For the purposes of this Article, this pressure to conform is called the “conformity problem.” The conformity problem is a phenomenon that has been documented in cognitive psychology experiments and demonstrated in economic and social science modeling.²⁴ This Article hypothesizes that this conformity problem is a significant factor in judicial decision-making around forensic science.

In this context, the problem goes beyond the requirements of precedent—as judges view decision after decision admitting forensic science evidence (the vast majority of which are nonbinding), the judges conform their decisions and similarly admit the evidence. The conformity problem encompasses a complicated interrelationship among a desire to conform, personally held beliefs, and those who seek to influence judicial decisions.

This conformity problem, however, is not irreversible. Some judges’ personal characteristics may make them particularly open to forensic science challenges and willing to deviate from the norm. Once those judges decide to limit or exclude the testimony, they open the door for other judges to deviate and eventually a new norm can be established.²⁵ As groups of judges become more critical of forensic science evidence, these groups can influence other groups, then the

²³ *Id.*

²⁴ See discussion *infra* Sections III.B.1 and III.B.2, and Part IV.

²⁵ See discussion *infra* note 226.

conformity problem will push judges to be critical of forensic science evidence. Examining judicial decision-making from the perspective of the conformity problem is important because it provides a path toward ensuring that only reliable forensic science evidence goes before juries.

The issue with judicial decision-making regarding forensic science in criminal cases is not that judges allow the evidence in when the evidence should always be excluded. The solution is not the exclusion of all forensic science. The issue is that most judges in criminal cases are not sufficiently scrutinizing this evidence or ensuring its reliability.

This Article examines the problem of judicial conformity through the lens of a particular forensic science: firearm and toolmark examination (FTE). This is for two reasons. First, it permits comparison over a broad range of judicial decisions on a specific issue. Second, FTE is an interesting area of forensic science because it is neither clearly reliable (like DNA evidence) nor clearly unreliable (like bitemark or arson evidence). FTE has the potential to provide useful information to juries and judges, but studies have yet to establish the extent to which the information is useful.

Part II provides context for the conditions in which judges make decisions about the admissibility and reliability of forensic science evidence, including summaries of government-sponsored reports and developments since 2008, a discussion of the *Frye* and *Daubert* standards as applied to these issues, and an introduction to FTE. Part III provides the theoretical framework for the conformity problem and explores how it connects with judicial decision-making and forensic science. Part IV then examines case studies of judicial decision-making that could provide evidence of the conformity problem. This Part starts with an examination of *State v. Henderson*,²⁶ a New Jersey case on eyewitness identification that resulted in a nationwide shift in how courts and advocates address this evidence. It will then discuss a series of cases in the District of Columbia in which the law on FTE is slowly changing and two cases out of Massachusetts on FTE that have had a national, but not local, impact. Finally, the Conclusion asserts that once a single court or series of courts changes its approach to forensic science, other courts will follow.

²⁶ 27 A.3d 872 (N.J. 2011).

II. HOW *DAUBERT* AND *FRYE* HAVE FAILED TO KEEP OUT FAULTY FORENSIC SCIENCE DESPITE EVIDENCE OF FLAWS

Though the word “science” is in the name, science has played a very small role in the development of most forensic sciences. Some do have roots in science, such as medicolegal death investigation and DNA.²⁷

[But] [t]he evolution of other forensic disciplines, particularly those related to pattern evidence, followed a different course, having been developed primarily within law enforcement environments or at the behest of law enforcement. Disciplines such as fingerprints, firearms and tool marks, blood stain pattern analysis, tread impression analysis, and bite mark analysis matured largely outside of the traditional scientific community during a time when admissibility standards for scientific evidence had yet to be formulated. Thus, admissibility of such evidence rightly or wrongly created judicial precedent in decisions that often did not, or could not, involve the level of research that would today be needed to establish scientific validity.²⁸

Neither the rigors of the scientific method nor a strict admissibility standard ensured that pattern-matching areas of forensic science were scientifically sound. In 2009, the NAS Report concluded that “[w]ith the exception of nuclear DNA analysis, . . . no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.”²⁹ This report examined almost all areas of forensic science, from fingerprint comparison to bloodstain pattern analysis to forensic odontology (bitemark analysis) to FTE.³⁰

Since the early 2000s, scientific and governmental bodies have grown interested in better understanding the validity of forensic sciences and their use in court. Despite this interest and a growing body of evidence that forensic sciences can be faulty and unreliable, most

²⁷ See, e.g., Suzanne Bell et al., *A Call for More Science in Forensic Science*, 115 PROC. NAT'L ACAD. SCI. U.S. 4541, 4542 (2018).

²⁸ *Id.*; see also NAS REPORT, *supra* note 7, at 42 (“The fact is that many forensic tests . . . have never been exposed to stringent scientific scrutiny. Most of these techniques were developed in crime laboratories to aid in the investigation of evidence from a particular crime scene, and researching their limitations and foundations was never a top priority.”); Valena E. Beety, *Cops in Lab Coats and Forensics in the Courtroom*, 13 OHIO ST. J. CRIM. L. 543, 543–44 (2016) (“Forensic disciplines arose out of crime scene investigations and law enforcement’s search for compelling evidence to convict. In this unusual development, forensic results were not tested in a lab but rather in the field. Their reliability and importance were indicated by the rate of convictions for crimes, not by impartial scientific assessments.”).

²⁹ NAS REPORT, *supra* note 7, at 7.

³⁰ *Id.* at xvi.

courts have continued to admit this evidence without substantial inquiry, irrespective of whether a court applied the *Daubert* standard or the *Frye* standard.³¹ To this day, this evidence is still going before jurors, often without limitation.

This Part starts with an overview of select reports since 2008³² that identify substantial reliability issues with several areas of forensic science, specifically the National Research Council's Ballistic Imaging report (Ballistic Imaging Report),³³ the NAS Report,³⁴ and the President's Council of Advisors on Science and Technology's report on the scientific validity of feature-comparison methods (the "PCAST Report").³⁵ This Section is followed by a history of the *Daubert*³⁶ and *Frye*³⁷ decisions with a focus on how courts have evaluated forensic science within these legal frameworks. The last Section evaluates with more specificity the reliability issues with FTE.

A. Reports on the Reliability of Forensic Sciences Demonstrate Substantial Problems

1. Ballistic Imaging Report

In 2008, before the release of the comprehensive NAS Report, the National Research Council released a report on ballistic imaging. This report originated from a 2004 request from the National Institute of Justice (a division of the Department of Justice (DOJ)) for a report on "the issues raised by the computerized imaging ballistic technology,"³⁸ and specifically to "assess the feasibility, accuracy, and reliability, and technical capability of developing and using a national ballistic database as an aid to criminal investigations."³⁹

³¹ See generally Jessica G. Cino, *An Uncivil Action: Criminalizing Daubert in Procedure and Practice to Avoid Wrongful Convictions*, 119 W. VA. L. REV. 651 (2016) (discussing the role of forensic science evidence in wrongful convictions, the standards of admissibility of such evidence, and issues that arise out of these standards).

³² This list of studies and reports is not exhaustive. Other organizations have also studied these issues, including the American Association for the Advancement of Science, which has released two reports on forensic science disciplines: one on latent fingerprints and the other on fire investigation. See *Forensic Science Assessments: A Quality and Gap Analysis*, AM. ASS'N FOR THE ADVANCEMENT OF SCI., <https://www.aaas.org/resources/forensic-science-assessments-quality-and-gap-analysis> (last visited Jan. 7, 2020).

³³ NAT'L RES. COUNCIL, BALLISTIC IMAGING (2008).

³⁴ NAS REPORT, *supra* note 7.

³⁵ PCAST REPORT, *supra* note 7.

³⁶ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

³⁷ *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

³⁸ NAT'L RES. COUNCIL, *supra* note 33, at 2.

³⁹ *Id.* at 1–2.

To conduct such an assessment, though, the committee writing this report had to understand (a) the mechanisms by which firearms leave markings on bullets and casings and (b) whether such markings are unique.⁴⁰ “Very early in its work the committee found that this question cannot now be definitively answered.”⁴¹ The committee found that the markings left on bullets and casings were likely not completely random; thus, the committee had reason to believe that similar marks would likely exist on bullets and casings fired from the same firearm.⁴² The committee also found, however, that “the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.”⁴³

The committee noted some of the challenges of FTE, including that the characteristics of a single firearm can change over time due to use and corrosion,⁴⁴ bullets and casings can have markings that the firearm did not cause,⁴⁵ and problems with the tool that most FTE examiners use to make comparisons.⁴⁶ Compounding these problems, the committee noted that pressures on an FTE examiner to provide testimony helpful to the prosecutor resulted in FTE examiners expressing “their findings in bold absolutes—matches made to the same gun, to the exclusion of all other firearms in the world.”⁴⁷ These statements, however, were not based on any studies or objective standards. This report was the first comprehensive analysis of FTE and noted substantial issues with the practice.

2. NAS Report

In 2009, shortly after the release of the Ballistics Imaging Report, the National Academy of Sciences released a report titled *Strengthening Forensic Science in the United States: A Path Forward*.⁴⁸ This report arose from a 2006 federal statute directing an examination of the needs of the

⁴⁰ See *id.* at 3.

⁴¹ *Id.*

⁴² See *id.* at 81.

⁴³ *Id.*

⁴⁴ See NAT'L RES. COUNCIL, *supra* note 33, 76–78.

⁴⁵ See *id.* at 76–77 (discussing other marks that can be found on a bullet as a result of the “nonpristine nature of crime scene evidence”).

⁴⁶ *Id.* at 63. Many FTE examiners use a comparison microscope, which allows the examiner to simultaneously view the bullets or casings she is comparing. *Id.* Intuitively, one might assume that this would be a very helpful tool, but the comparison microscope is not as high quality a microscope as others, which can result in problems seeing the markings on the casings. *Id.* at 63–64.

⁴⁷ *Id.* at 67.

⁴⁸ NAS REPORT, *supra* note 7.

forensic science community.⁴⁹ In the three years between the creation of the committee working on the report and the issuance of the report, the committee heard a clear and consistent message:

The forensic science system, encompassing both research and practice, has serious problems that can only be addressed by a national commitment to overhaul the current structure that supports the forensic science community in this country. This can only be done with effective leadership at the highest levels of both federal and state governments, pursuant to national standards, and with a significant infusion of federal funds.⁵⁰

The report addressed not only the areas of forensic science that need additional research but also structural issues facing the field, such as a lack of governance.⁵¹

Though a substantial part of the report focuses on the need for further scientific study of many forensic science fields, the report also addresses how problems can occur due to human error or fraud, including contamination or mislabeling; falsifying results or misinterpreting the evidence; failing to provide exculpatory information to the defense; or exaggerating results on the witness stand.⁵² Thus, even the most studied and substantiated areas of science can result in wrongful convictions without certain structural safeguards.

In its evaluation of the various areas of forensic science, the NAS Report concluded the following: (a) DNA testing and laboratory chemical testing (such as testing used to determine whether a substance is a controlled dangerous substance) are scientifically sound;⁵³ (b) though some scientific evidence indicates that fingerprints are unique to each individual, the evidence has failed to demonstrate that fingerprints can be sourced to a particular individual or distinguished among several individuals;⁵⁴ (c) because the methods for analyzing

⁴⁹ *Id.* at 1.

⁵⁰ *Id.* at xx. This conclusion was based on testimony the committee heard from stakeholders in the forensic science field including managers of laboratories, teachers, scholars, and members of the legal profession and law enforcement. *Id.*

⁵¹ *Id.* at 77.

⁵² *Id.* at 45.

⁵³ *Id.* at 133, 135–36.

⁵⁴ NAS REPORT, *supra* note 7, at 142–45. This issue famously came to light when an FBI fingerprint expert incorrectly concluded that the fingerprints recovered from a 2004 Madrid bombing belonged to Brandon Mayfield, a Portland, Oregon resident. Jennifer Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 BROOK. L. REV. 1209, 1228–29 (2010). Three separate FBI fingerprint examiners concluded that Mr. Mayfield was the source before Spanish authorities convinced the FBI that the actual source was another man with remarkably similar fingerprints to Mr. Mayfield. *Id.* at 1229.

firearm toolmarks are not precisely defined, studies are necessary to establish that firearms leave unique markings;⁵⁵ (d) hair microscopy (the matching of a particular hair with a particular individual), without accompanying DNA analysis, is not scientifically supported;⁵⁶ (e) though it may be possible to individualize fibers and paints,⁵⁷ no studies exist to date and the only reliable conclusion regarding comparisons of fibers and paints is that the fibers or paint could have come from the same type of item;⁵⁸ (f) handwriting analysis may have value, though only limited research has been conducted;⁵⁹ (g) assumptions that arson investigators made in the past have been shown to be incorrect, and more study is required;⁶⁰ (h) though forensic odontology (bitemark analysis) can be used to exclude an individual, no evidence establishes that it can individualize;⁶¹ and (i) “the uncertainties associated with bloodstain pattern analysis are enormous,” but whether blood spattered quickly or slowly is determinable.⁶² What the *NAS Report* makes clear is that, other than DNA evidence and laboratory chemical analysis, substantial research is necessary to establish whether other areas of forensic science can actually do what experts have been claiming they can do for decades.

3. PCAST Report and the National Commission on Forensic Science

In 2016, the President’s Council of Advisors on Science and Technology released a report on forensic science based on a review of more than 2000 papers and in consultation with judges, statisticians, FBI scientists, other forensic scientists and practitioners, prosecutors, defense attorneys, academics, advocates, and agency representatives.⁶³ The report primarily examined six areas of feature-comparison forensic science:⁶⁴ (1) DNA analysis of single-source and simple-mixture

⁵⁵ NAS REPORT, *supra* note 7, at 154–55.

⁵⁶ *Id.* at 160–61.

⁵⁷ To individualize means that a particular sample recovered from a crime scene can be conclusively matched to a single source: essentially, that a source has “individual” characteristics such that any item from that source can be traced directly back to it. *See supra* note 8.

⁵⁸ NAS REPORT, *supra* note 7, at 162–63, 170.

⁵⁹ *Id.* at 166–67.

⁶⁰ *Id.* at 172–73.

⁶¹ *Id.* at 176.

⁶² *Id.* at 178–79.

⁶³ PCAST REPORT, *supra* note 7, at 17–19.

⁶⁴ The PCAST REPORT defines “feature-comparison” forensic science as “methods that attempt to determine whether an evidentiary sample (e.g., from a crime scene) is or is not associated with a potential ‘source’ sample (e.g., from a suspect), based on the

samples,⁶⁵ (2) DNA analysis of complex-mixture samples,⁶⁶ (3) bitemarks, (4) latent fingerprints, (5) firearms identification, and (6) footwear analysis.⁶⁷ The PCAST Report is largely responsive to the NAS Report, focusing on the issue that had not been addressed since the release of the NAS Report: whether forensic sciences have “fundamental scientific validity and reliability.”⁶⁸

In evaluating the six areas of forensic science, the PCAST Report made the following findings: (1) DNA single-source and simple-mixture analysis is foundationally valid, though errors can occur in particular instances due to practitioner error;⁶⁹ (2) complex DNA analysis is not foundationally valid and more empirical studies need to be conducted to evaluate the validity of the forensic science;⁷⁰ (3) bitemark analysis is not foundationally valid and the problems with this area of forensic science are so substantial that they are likely insurmountable;⁷¹ (4) latent fingerprint analysis is foundationally valid but has a substantial false positive rate and issues with application because of the subjectivity of the practice;⁷² (5) FTE is not presently foundationally valid, but with further study could be, though issues will likely persist because of the error rates and lack of adequate proficiency testing;⁷³ and (6) footwear analysis is not foundationally valid and no empirical evidence supports the conclusions it purports to make.⁷⁴ The report recommended the judiciary more substantially consider the scientific methods of the forensic sciences and limit the degree of certainty to which experts can testify.⁷⁵

presence of similar patterns, impressions, or other features in the sample and the source.” *Id.* at 17–18. This Article refers to such areas of forensic science as “pattern-matching.”

⁶⁵ A single-source DNA sample is one that includes DNA from only one individual. *Id.* at 70. A simple-mixture sample is one where the DNA might come from two individuals, but one of the contributors is a known contributor. *Id.* Therefore, identifying the DNA profile of the unknown individual is “simple” because one of the two profiles is known.

⁶⁶ A complex mixture is one where the DNA comes from multiple unknown individuals. *Id.* at 75.

⁶⁷ *Id.* at 7.

⁶⁸ *Id.* at 39.

⁶⁹ PCAST REPORT, *supra* note 7, at 69. See also *supra* note 65 for an explanation of single source and simple-mixture DNA profiles.

⁷⁰ *Id.* at 82. See also *supra* note 66 for an explanation of complex-mixture DNA profiles.

⁷¹ *Id.* at 87.

⁷² *Id.* at 95–96.

⁷³ *Id.* at 111–12.

⁷⁴ *Id.* at 117.

⁷⁵ PCAST REPORT, *supra* note 7, at 19.

In 2013, President Obama created the National Commission on Forensic Science, which included commissioners from varied backgrounds, including scientists, prosecutors, defense attorneys, judges, and practitioners.⁷⁶ The Commission was created to recommend “strategies for enhancing quality assurance in forensic science units.”⁷⁷ Former U.S. Attorney General Jeff Sessions disbanded this Commission in 2017.⁷⁸

B. Daubert, Frye, and Forensic Science

To evaluate whether judges are appropriately admitting forensic science in the courtroom, one must first understand the legal standards that should be guiding these decisions: *Daubert* and *Frye*. Though federal courts and many state courts have adopted the *Daubert* standard, *Frye* is still used in some state courts, particularly in the criminal context.⁷⁹

1. *Frye*

The Court of Appeals of the District of Columbia decided *Frye v. United States*⁸⁰ in 1923, resolving the novel issue of whether expert testimony on the systolic blood pressure deception test (essentially a lie detector test) was admissible at trial.⁸¹ The court held that “when the question involved does not lie within the range of common experience or common knowledge, but requires special experience or special knowledge, then the opinions of witnesses skilled in that particular science, art, or trade to which the questions relates are admissible in evidence.”⁸² The court drew a distinction between testimony based on science that was demonstrable rather than merely experimental. Noting that it was difficult to draw a line between the two, however, the court held that to be admissible, the scientific principle or discovery “must be sufficiently established to have gained general acceptance in the particular field to which it belongs.”⁸³

⁷⁶ Jules Epstein, *The National Commission on Forensic Science: Impactful or Ineffectual?*, 48 SETON HALL L. REV. 743, 748–49 (2018).

⁷⁷ *Id.* at 748.

⁷⁸ *Id.* at 743.

⁷⁹ Cino, *supra* note 31, at 681–85. Despite the liberty interest at stake in criminal cases, judges are more stringent in admitting expert testimony in civil cases than criminal cases, such as by applying different *Daubert* standards in civil cases than criminal cases. *Id.* at 685–86.

⁸⁰ 293 F. 1013 (D.C. 1923).

⁸¹ *Id.* at 1013.

⁸² *Id.* at 1014.

⁸³ *Id.*

In jurisdictions that apply *Frye* to forensic science, challenging the admissibility of the evidence is particularly difficult. Because courts have routinely admitted forensic science evidence in the past, “general acceptance” is an easy standard to meet.⁸⁴ In the context of FTE, some courts find that *Frye* does not apply at all because FTE is not new or novel;⁸⁵ other courts rely on precedent that FTE is generally accepted and therefore admissible under *Frye*.⁸⁶

2. *Daubert*

Frye persisted as the dominant standard for seventy years until the Supreme Court decided *Daubert v. Merrell Dow Pharmaceuticals, Inc.*⁸⁷ In *Daubert*, a case regarding the impact of an anti-nausea drug on birth defects, the Court held that revisions to the Federal Rules of Evidence displaced *Frye* and established the following test for the admissibility of specialized testimony: “whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or

⁸⁴ See JoAnne A. Epps & Kevin Todrow, *Refried Forensics: Screening Expert Testimony in Criminal Cases Through Frye Plus Reliability*, 48 SETON HALL L. REV. 1161, 1181–90 (2018) (discussing judicial recognition of general acceptance and jurisdictions that deny *Frye* hearings because of this judicial recognition); Paul A. Rodrigues, *Toward a New Standard for the Admission of Expert Evidence in Illinois: A Critique of the Frye General Acceptance Test and an Argument for the Adoption of Daubert*, 34 S. ILL. U. L.J. 289, 296–308 (2010) (finding that Illinois courts have relied on prior determinations of general acceptance, even where those determinations have been incorrect); Simon A. Cole, *Out of the Daubert Fire and into the Fryeing Pan? Self-Validation, Meta-Expertise and the Admissibility of Latent Print Evidence in Frye Jurisdictions*, 9 MINN. J.L. SCI. & TECH. 453, 539–40 (2008) (discussing general acceptance in the context of latent fingerprint evidence and how courts recognize general acceptance in the judicial community, therefore “turn[ing] a *Frye* analysis into an exercise in following legal precedent”).

⁸⁵ See, e.g., *State v. Goudeau*, 372 P.3d 945, 982 (Ariz. 2016) (holding that a *Frye* hearing was not required for testimony that every bullet and casing retrieved was fired from the same gun because it was not a new form of testimony and Arizona courts had previously upheld its admissibility.); *King v. State*, 89 So. 3d 209, 228–29 (Fla. 2012) (holding that FTE is not new or novel, the court noted that “tool-mark identification in the context of ballistics has been used in the criminal context since at least 1929, and in Florida since at least 1937”); *People v. Jones*, 34 N.E.3d 1065, 1077 (Ill. App. Ct. 2015) (finding that defendant could not challenge expert testimony under *Frye* because “[f]irearm and toolmark identification is not new or novel”); *Commonwealth v. Whitacre*, 878 A.2d 96, 101 (Pa. Super. Ct. 2005) (“The comparison microscope technique has been in use since 1930’s and is an accepted methodology by the Association of Firearms and Toolmark Examiners . . . it is neither new nor original, rather the sort that is offered all the time.”).

⁸⁶ See, e.g., *People v. Robinson*, 2 N.E.3d 383, 402 (Ill. Ct. App. 2013) (holding that judicial decisions uniformly conclude toolmark and firearms identification is generally accepted and admissible at trial); *People v. Givens*, 912 N.Y.S. 2d 855, 857 (N.Y. Sup. Ct. 2010) (“This court was unable to find any cases where firearms and toolmark identification was found to be unreliable or no longer scientifically acceptable. Nor were there instances where the testimony was ruled to be inadmissible.”).

⁸⁷ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

determine a fact in issue.”⁸⁸ To determine if the expert satisfies the test, the court undertakes a “preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.”⁸⁹

The Court held that “the requirement that an expert’s testimony pertain to ‘scientific knowledge’ establishe[d] a standard of evidentiary reliability,”⁹⁰ whereas the need for the expert’s testimony to assist the trier of fact in understanding or determining a fact in issue went primarily to the relevance of that evidence.⁹¹ To assess reliability, the Court encouraged lower courts to consider whether the theory or technique can be and has been tested; whether the theory or technique has been subjected to peer review and publication; the known or potential error rate of a particular scientific technique; and the existence and maintenance of standards controlling the operation of a particular scientific technique.⁹² The Court noted that general acceptance can be relevant to the reliability inquiry; though not required, the assessment permits “explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.”⁹³

Daubert should create a higher bar for the admissibility of forensic science evidence than *Frye*, particularly where the reliability of the evidence is in question. In response to *Daubert*, courts have expressed a desire to avoid grandfathering in scientific principles that are no longer accepted,⁹⁴ and the courts’ gatekeeping function, in theory, encourages hearings on the admissibility of such evidence. In the context of FTE, for example, the hearings tend to relate to the reliability

⁸⁸ *Id.* at 592.

⁸⁹ *Id.* at 592–93.

⁹⁰ *Id.* at 590.

⁹¹ *Id.* at 591.

⁹² *Id.* at 592–94.

⁹³ *Id.* at 594. Following the *Daubert* decision, the Supreme Court issued two subsequent decisions that, combined with *Daubert*, are known as the *Daubert* trilogy. Moriarty, *supra* note 16, at 1694, n. 48. In *General Electric Co. v. Joiner*, the Supreme Court established abuse of discretion as the appropriate standard for an appellate court reviewing a trial court’s decision regarding the admissibility of evidence. 522 U.S. 136, 139 (1997). *Kumho Tire Co. v. Carmichael* established that a court’s basic gatekeeping obligation applies to all expert testimony, not just scientific testimony. 526 U.S. 137, 141 (1999).

⁹⁴ See, e.g., *United States v. Green*, 405 F. Supp. 2d 104, 123 (D. Mass. 2005) (“This reliance on long-standing use of ballistics evidence in the courts is troubling. It runs the risk of ‘grandfathering in irrationality,’ without reexamining it in the light of *Kumho* and *Daubert*. It arguably ignores the mandate of *Daubert*, especially where the courts are relying on pre-*Daubert* acceptance of a given scientific technique.”).

of the evidence. In one particular case, *United States v. Otero*, the United States District Court for the District of New Jersey conducted a three-day *Daubert* hearing on the defendant's motion to preclude the testimony of the government's FTE expert witness.⁹⁵ In addition to its analysis under the various factors, the court noted that FTE is generally accepted in the forensic community and that other courts have admitted the evidence, even when those courts criticized the practice and the studies upon which the government witness relied.⁹⁶ As the *Otero* case illustrates courts have been reluctant to change course when it comes to admitting forensic science evidence and have continued to grandfather in the evidence despite report after report urging courts to reevaluate this evidence.⁹⁷

C. The Limited Reliability of Firearm and Toolmark Examination

Pattern-matching forensic science⁹⁸ includes the areas of forensic science in which a trained individual compares one or more samples of items and makes a subjective determination as to whether the items have sufficient similarities such that they likely came from the same source (or whether they are sufficiently different that one can be excluded from having the same source).⁹⁹ Some examples of this area of forensic science are latent fingerprint examination, FTE, hair microscopy, and bitemark analysis.¹⁰⁰

The expert in pattern-matching forensic science is not relying on any specific and objective criteria in making these determinations.¹⁰¹ The individual evaluating the evidence looks at it, thinks about the areas of similarity and difference, and then reaches a conclusion based on her prior experiences, training, and observations of what these pieces of evidence look like.¹⁰² The following rudimentary comparison helps illustrate why this is problematic: high school math students are told to "show their work" on homework and tests so a teacher can tell if the student got the right answer because of her work, chance, or cheating. On the other hand, if the student got the wrong answer, the teacher can tell from her work whether that is because she made a small miscalculation somewhere or if she simply does not understand the

⁹⁵ 849 F. Supp. 2d 425 (D.N.J. 2012).

⁹⁶ *Id.* at 435.

⁹⁷ See discussion *infra* Section III.B.

⁹⁸ The *PCAST Report* refers to pattern-matching forensic science as "forensic feature-comparison methods." *PCAST REPORT*, *supra* note 7, at 1; see *supra* note 64.

⁹⁹ *PCAST REPORT*, *supra* note 7, at 23.

¹⁰⁰ *Id.*

¹⁰¹ See *NAS REPORT*, *supra* note 7, at 141.

¹⁰² See *PCAST REPORT*, *supra* note 7, at 48–50.

concept. This allows the teacher to check the individual's work—to make sure it is right for the right reasons and, if wrong, help fix the problem for the future. The problem with these pattern-matching evaluations is that there is no objective way to evaluate whether the expert made the correct subjective determination or how she reached that determination.¹⁰³

The NAS Report and PCAST Report have found (or failed to find) varying degrees of support for pattern-matching disciplines. Though bitemark analysis and hair microscopy have been widely discredited,¹⁰⁴ latent fingerprint evidence and FTE are among those that the NAS Report and PCAST Report acknowledge could have evidentiary support, should they be subjected to more studies and regulation.¹⁰⁵

FTE is premised on the idea that each firearm will leave individual and unique markings on the cartridge casings as they leave the gun.¹⁰⁶ These markings are allegedly unique because of the way that firearms are manufactured—as each firearm is produced, the tool that makes the firearm is dulled because metal is cutting metal; therefore, each firearm has markings slightly different than the firearm made before it.¹⁰⁷ Then, when this firearm is fired, those features are imprinted into the casing, which has expanded and scraped against the inside of the firearm as it was fired.¹⁰⁸ Other parts of the firearm, such as a spring called an “extractor,” can also leave marks on the casing.¹⁰⁹

When an FTE examiner obtains bullets or casings for examination, the current best-case scenario is that they follow the Association of Firearm and Tool Mark Examiners (AFTE) guidelines for examination.¹¹⁰ The AFTE specifies that an examiner should examine the casings using vision-enhancing tools, generally a microscope, to

¹⁰³ The inability to easily check an examiner's work can lead to widespread error. In 2008, Detroit shuttered its police lab after an audit demonstrated a “shocking and appalling” degree of error. Bunkley, *supra* note 2. Auditors found a likely 10% error rate in FTE—the errors were false positives, meaning that people were likely wrongfully convicted on this evidence. *Id.* Again, judges have continued to admit this evidence without scrutiny even after scandals like that in Detroit and reports like the 2009 NAS REPORT and PCAST REPORT.

¹⁰⁴ See generally Saks & Koehler, *supra* note 8, at 892.

¹⁰⁵ PCAST REPORT, *supra* note 7, at 17–19.

¹⁰⁶ United States v. Green, 405 F. Supp. 2d 105, 110 (D. Mass. 2005). The intricate details of how firearms leave markings on bullets and casings are discussed in far greater detail in the *Ballistics Imaging Report*. NAT'L RES. COUNCIL, *supra* note 33.

¹⁰⁷ Green, 405 F. Supp. 2d at 110.

¹⁰⁸ *Id.*

¹⁰⁹ *Id.*

¹¹⁰ See United States v. Monteiro, 407 F. Supp. 2d 351, 374 (D. Mass. 2006) (holding that the FTE expert could not testify because his methods were not consistent with the AFTE guidelines).

identify class, subclass, and individualized characteristics.¹¹¹ Class characteristics are those that would identify the casings as having been fired by a particular type of firearm, such as a Glock, and individualized characteristics would be such that an examiner could say the casing comes from one specific Glock.¹¹² Subclass characteristics are those that fall somewhere between class and individual characteristics, which may be unique to a particular set of firearms, but not an entire set of firearms.¹¹³

However, studies have not substantiated that a firearm has individualized characteristics such that casings could have come from one and only one firearm rather than the firearm manufactured immediately before or after it or at some other time.¹¹⁴ The problem is that how an examiner decides what is a class characteristic versus a subclass characteristic versus an individual characteristic is neither standardized nor objective.¹¹⁵ The AFTE does not delineate the number of class, subclass, or individualized characteristics that the items being compared must have in common or must be different before the examiner can call it a match or assert exclusion.¹¹⁶

Rather, under the AFTE method, the examiner may opine that a specific tool or firearm caused certain toolmarks when significant agreement exists between the two sets of marks.¹¹⁷ The AFTE goes on to define agreement as significant “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

¹¹¹ *Summary of the Examination Method*, ASS’N OF FIREARM & TOOL MARK EXAM’RS, <https://afte.org/resources/swggun-ark/summary-of-the-examination-method> (last visited Feb. 5, 2020).

¹¹² *United States v. Johnson*, No. 16 Cr. 281, 2019 WL 1130258, at *7–8 (S.D.N.Y. 2019).

¹¹³ NAT’L RES. COUNCIL, *supra* note 33, at 58.

¹¹⁴ PCAST REPORT, *supra* note 7, at 17–19. Part of the problem in determining individualization is that firearms are often manufactured by manufacturing parts of the firearm separately, then those parts are assembled, but not in order of when the individual parts were manufactured. NAT’L RES. COUNCIL, *supra* note 33, at 49. Therefore, even though two firearms might come off the assembly line at the same time, their individual parts could have been manufactured at varying times.

¹¹⁵ *See* NAS REPORT, *supra* note 7, at 141.

¹¹⁶ NAT’L RES. COUNCIL, *supra* note 33, at 59–60 (describing the AFTE Theory of Identification, which does not include a requirement that a certain number of class, subclass, or individual characteristics must be identified in order to reach a particular conclusion).

¹¹⁷ NAS REPORT, *supra* note 7, at 153.

been produced by the same tool.”¹¹⁸ As the PCAST Report notes, the theory of AFTE examinations is “circular.”¹¹⁹ Essentially, and as AFTE concedes, the conclusions of these experts are subjective.¹²⁰

Three problems continue to persist regarding FTE: (1) no study has established the central premise upon which FTE is based—that the toolmarks on these casings are unique or the degree to which they are unique;¹²¹ (2) a dearth of studies establishing an error rate;¹²² and (3) FTE examiners have not been universally subject to proficiency testing, nor has the available proficiency testing been sufficiently rigorous to actually determine whether an examiner is reliable.¹²³ That being said, even those critical of FTE admit that studies may eventually demonstrate that this area of forensic science is reliable to some degree—the problem is that those studies do not yet exist.¹²⁴

Despite these problems, which were identified more than a decade ago and remain unresolved, judges in courts across the country not only continue to admit this evidence but often do so without critical analysis. This is the case even when defense attorneys raise these concerns either in motions for *Frye* or *Daubert* hearings or during such a hearing.¹²⁵ The question again becomes, why? Why, in the face of so much evidence that forensic science must be rigorously tested before being admitted, are judges unwilling to put forensic science to the test?

¹¹⁸ *AFTE Theory of Identification as it Relates to Toolmarks*, ASS’N OF FIREARM & TOOL MARK EXAM’RS, <https://afte.org/about-us/what-is-afte/afte-theory-of-identification> (last visited Jan. 20, 2020).

¹¹⁹ PCAST REPORT, *supra* note 7, at 17–19.

¹²⁰ *AFTE Theory of Identification as it Relates to Toolmarks*, *supra* note 118. According to the AFTE theory of identification, “the interpretation of individualization/identification is subjective in nature.” *Id.*; see also PCAST REPORT, *supra* note 7, at 17–19.

¹²¹ PCAST REPORT, *supra* note 7, at 11.

¹²² *Id.* at 17–19. In an extensive review of literature prior to 2009, the PCAST REPORT found no studies that accurately assessed the reliability of FTE. At the time of the report’s publication, only one “appropriately designed black-box study” had been completed since 2009 on the reliability of FTE, which it considered insufficient to establish scientific validity and reliability. *Id.* at 11.

¹²³ See *Monteiro*, 407 F. Supp. 2d at 367; Adina Schwartz, *A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification*, 6 COLUM. SCI. & TECH. L. REV. 2, 26–27 (2005) (critiquing the use of declared rather than blind proficiency testing, and reports that laboratories spend more time obtaining results in declared proficiency tests than actual casework).

¹²⁴ PCAST REPORT, *supra* note 7, at 17–19.

¹²⁵ See *supra* notes 86, 93 and accompanying text (discussing cases where judges have admitted FTE based on prior judicial decisions); see also *United States v. Ashburn*, 88 F. Supp. 3d 239, 244, 247 (E.D.N.Y. 2015) (denying defense request for a *Daubert* hearing and, without independent analysis, finding that FTE evidence satisfies each *Daubert* factor based on prior judicial decisions).

III. JUDICIAL DECISION-MAKING, FORENSIC SCIENCE, AND THE CONFORMITY PROBLEM

Given the critiques in the Ballistic Imaging Report, NAS Report, and PCAST Report, combined with the increased gatekeeping power provided under *Daubert*, one might expect to find judges substantially limiting the use of forensic science evidence. This, however, has not been the case. Cognitive biases, pressures to conform, false beliefs, and outside influence are surely impacting judges' abilities to respond rationally to the new information about problems with forensic science.

Judges are people like the rest of us: subject to the same biases and motivations that influence our abilities to make informed and correct decisions.¹²⁶ "They are swayed by heuristic decision-making, friendships, beauty, the strength of a case, public opinion, fear of reversal, and the normal set of cognitive biases to which we all are subject: expectation bias, hindsight bias, confirmation bias, tunnel vision, and so forth."¹²⁷ Scholars from psychologists to law professors to economists have studied judges and how they make decisions.¹²⁸ In many ways, the results are reassuring: judges tend to evaluate problems just like everyone else. But that conclusion can also be problematic—judges are tasked with making complex decisions that can substantially impact people's lives yet are often using problematic decision-making shortcuts.

Judicial decision-making is primarily a combination of internal processing and external influences. Internal processing is studied largely through cognitive psychology. Many studies and reviews of judicial decision-making have focused on cognitive psychology—how the brain processes certain information when making decisions.¹²⁹

¹²⁶ Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 784 (2001) (citing JEROME FRANK, *COURTS ON TRIAL: MYTH AND REALITY IN AMERICAN JUSTICE* 410 (1949)).

¹²⁷ Jane Campbell Moriarty, *Will History Be Servitude: The NAS Report of Forensic Science and the Role of the Judiciary*, 2010 UTAH L. REV. 299, 317–18 (2010).

¹²⁸ Jeffery T. Ulmer & Brian D. Johnson, *Organizational Conformity and Punishment: Federal Court Communities and Judge-Initiated Guidelines Departures*, 107 J. CRIM. L. & CRIMINOLOGY 253 (2017); Chris Guthrie et al., *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1 (2007); Sophie Harnay & Alain Marciano, *Judicial Conformity Versus Dissidence: An Economic Analysis of Judicial Precedent*, 23 INT'L REV. L. & ECON. 405 (2004); Guthrie et al., *supra* note 126, at 777.

¹²⁹ See *Cognitive Psychology*, A.P.A. DICTIONARY OF PSYCHOL., <https://dictionary.apa.org/cognitive-psychology> (last visited Nov. 11, 2020) (defining cognitive psychology generally as a "branch of psychology that explores the operation of mental processes related to perceiving, attending, thinking, language, and memory, mainly through inferences from behavior"). See generally John N. Drobak & Douglass C. North, *Understanding Judicial Decision-Making: The Importance of Constraints on Non-Rational Deliberations*, 26 WASH. U. J.L. & POL'Y 131 (2008); Guthrie et al., *supra* note 126; Guthrie

Without understanding the impact of external influences on internal processing, however, only part of the picture is painted. Similarly, without understanding cognitive psychology, how external influences can have such a substantial impact on decision-making will remain elusive. Thus, though this Article focuses on the external influences, understanding the internal processing through cognitive psychology is essential background for any evaluation of judicial decision-making.

This Part first discusses the cognitive psychology research connected with judicial decision-making. Against this backdrop is a discussion of the conformity problem and forensic science: the theoretical foundation of it and how it may be influencing judges to admit faulty forensic science in criminal trials without critical evaluation.

A. *The Impact of Internal Cognitive Biases on Judicial Decision-Making*

Judges rely on cognitive biases, sometimes called cognitive illusions or heuristics, in making decisions—these are shortcuts that allow people to take information, quickly analyze it, make a few guesses (the shortcuts), and then reach a conclusion.¹³⁰ For example, with the cognitive bias “anchoring,” if a person is asked to guess how much a pencil costs and is told that the pencil costs less than \$10,000, the person is likely to guess a higher number than a person not told about the \$10,000 limit, even though it is preposterous that a pencil would cost even close to \$10,000.¹³¹ People use that “anchor” of \$10,000 as a shortcut to try to determine the cost of the pencil, and some might, perhaps unconsciously, assume that if \$10,000 is mentioned, the pencil must be worth more than they otherwise would have thought.¹³²

In *Inside the Judicial Mind*, researchers surveyed 167 federal magistrate judges across five common cognitive biases: anchoring, framing, hindsight bias, representativeness bias, and egocentric bias.¹³³ As described above, anchoring is when one makes “estimates based on irrelevant starting points.”¹³⁴ Framing is when one treats “economically equivalent gains and losses differently,” such as a person preferring to

et al., *supra* note 128; Dan Simon, *A Psychological Model of Judicial Decision Making*, 30 Rutgers L.J. 1 (1998).

¹³⁰ Guthrie et al., *supra* note 126, at 780 (citing Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 Sci. 1124, 1124 (1974)).

¹³¹ *Id.* at 789.

¹³² *Id.* at 788.

¹³³ *Id.* at 778.

¹³⁴ *Id.* at 784.

buy something with a 90% chance of success rather than a 10% chance of failure—even though they are the same thing.¹³⁵ Hindsight bias is when one perceives “past events to have been more predictable than they actually were,” such as believing that passing the California Bar Examination was an inevitability.¹³⁶ The representative bias is when one ignores “important background statistical information in favor of individuating information.”¹³⁷ For example, people often have a greater fear of a plane crashing than a car crashing, even though the latter is more likely because plane crashes are covered far more significantly in the news. Finally, the egocentric bias is when one overestimates “one’s own abilities,” including whether one knows what the right outcome of a decision should be.¹³⁸

The researchers found that judges are susceptible to all five cognitive biases.¹³⁹ In fact, judges are as susceptible to three of the five biases as the general population (anchoring, hindsight bias, and egocentric bias).¹⁴⁰ They are less susceptible, while still being susceptible, to framing and the representativeness bias.¹⁴¹ Therefore, when judges are making decisions, they are likely using the same cognitive shortcuts that the general population uses in daily life.

Another study from the same researchers found that judges primarily rely on intuition when making decisions, but such intuition could be overridden if the judges had a chance for subsequent deliberation.¹⁴² The researchers term this the “intuitive-override” process of decision-making.¹⁴³ The researchers conclude that while intuitive decision-making can lead to accurate results,¹⁴⁴ reliance on intuition can be particularly problematic in the legal system where “intuition is . . . the likely pathway by which undesirable influences, like the race, gender, or attractiveness of parties, affect the legal system.”¹⁴⁵

¹³⁵ Guthrie et al., *supra* note 126, at 784.

¹³⁶ *Id.*

¹³⁷ *Id.*

¹³⁸ *Id.*

¹³⁹ *Id.* at 787.

¹⁴⁰ *Id.* at 816.

¹⁴¹ *Id.*

¹⁴² Guthrie et al., *supra* note 128, at 29.

¹⁴³ *Id.* at 3.

¹⁴⁴ *Id.* at 5.

¹⁴⁵ *Id.* at 31. The ability to make quick decisions, some of which can be accurate, was famously examined in Malcolm Gladwell’s book, *BLINK*. MALCOLM GLADWELL, *BLINK: THE POWER OF THINKING WITHOUT THINKING* (2005). This book examined our ability to make accurate, quick decisions, as well as why those decisions are sometimes wrong and how people can master making such decisions accurately. See generally *id.*

Scholars have also applied judicial reliance on cognitive biases to decision-making in forensic science. Professor Jane Campbell Moriarty argues that judges view decisions regarding admission of forensic science evidence as simple.¹⁴⁶ Once a judge perceives the decision as simple, she is more likely to rely on the cognitive shortcuts described in the *Inside the Judicial Mind* study, which can lead to erroneous results.¹⁴⁷ Most forensic science determinations, however, are not easy—the method of examination, history of admission, and research in the areas are complex and require considerable attention.¹⁴⁸ That judges view this evidence as “easy” to admit can be understandable given that prosecutors often frame the issue as, “this evidence has been admitted in virtually all courts for decades and has properly formed the basis of many convictions.”¹⁴⁹ In addition, because many judges come from prosecutors’ offices, they may believe they have a firm grasp of criminal evidentiary issues, trust the prosecutors making these assertions, or have previously relied on the evidence when they were prosecutors.¹⁵⁰ Thus, they believe they can make more intuitive decisions.

¹⁴⁶ Moriarty, *supra* note 16, at 1708 (“Courts have categorized FCM evidence admissibility as a simple problem and have avoided addressing the serious questions about its reliability.”). In fact, in a District of Columbia Superior Court case, *United States v. Tibbs*, the judge presiding over the *Daubert* hearing noted,

This discipline and the dispute surrounding it is more complex than it initially seems. Full exploration of the issues surrounding the reliability of this evidence for at least a somewhat experienced judge to understand it required several days of testimony from multiple expert witnesses, close evaluation of numerous applied science studies, and some education or understanding of issues related to study design, statistics, and methodology, as well as advocacy by counsel on both sides, who are specially tasked by their offices to deal with forensic science issues.

When I say it’s more complex than it initially seemed, before we had this hearing I thought this was something that we could deal with in an afternoon.

Tr. of *Daubert* Hr’g., at 60, *United States v. Tibbs*, No. 2016-CF1-19431, 2019 D.C. Super. LEXIS 9 (Sept. 5, 2019).

¹⁴⁷ See Moriarty, *supra* note 16, at 1704.

¹⁴⁸ See, e.g., *United States v. Willock*, 696 F. Supp. 2d 536 (D. Md. 2010); Tr. of *Daubert* Hr’g., *supra* note 146 at 60. In *Willock*, the magistrate judge who presided over the hearing on the admissibility of FTE testimony and the degree of the certainty to which the expert could testify not only reviewed several expert affidavits but also reviewed “many published studies, journal articles, and cases” in rendering his thorough report. 696 F. Supp. 2d at 572.

¹⁴⁹ Moriarty, *supra* note 16, at 1704.

¹⁵⁰ See Clark Neily, *Are a Disproportionate Number of Federal Judges Former Government Advocates?*, CATO INST. (Sept. 18, 2019), <https://www.cato.org/publications/studies/are-disproportionate-number-federal-judges-former-government-advocates> (demonstrating that former prosecutors are appointed to the federal bench at a rate of 4:1 when compared with former defense attorneys).

These studies and analyses on the cognitive factors that influence judicial decision-making clearly demonstrate that judges are as likely or almost as likely as the rest of us to rely on cognitive biases, including when making decisions about forensic science. Though these perspectives help explain part of the puzzle, judges are not ruling in vacuums—it is not just their individual cognition that impacts how judges rule. Judges make decisions in the context of the criminal legal system—a community consisting of other judges, prosecutors, defense attorneys, police officers, and other state and institutional actors. The ties among these community members and their influence on each other is a substantial piece of the decision-making process.

B. The Relationship Among External Conformity Pressures, Judicial Decision-Making, and Forensic Science

Just as we all often rely on cognitive biases, we can also fall under the spell of conformity and the pressures to adhere to a belief even in the face of conclusive evidence to the contrary. This is the conformity problem. From an evolutionary and practical standpoint, this makes sense: as we move through the world, our experiences lead us to believe that some things operate in a particular way. This could be anything from a belief that money does not actually grow on trees or that purple cows do not fly through the sky.¹⁵¹ If someone comes to us and says, “Money grows on trees,” or “I just saw a purple cow flying through the sky,” we would rightly believe that the person was using hyperbole, kidding, or delusional.¹⁵²

This knowledge we gain is not only based on our own experiences but also on what we have learned from people around us whom we trust to provide us with accurate information. Most everyone believes the earth moves around the sun—a belief that is not based on direct experience but based on what other people have told us.¹⁵³ Because we trust those in our community, members of the same community tend to have beliefs that coalesce around each other, creating norms. Many judges likely believe that pattern-matching forensic science is reliable based either on personal experience or what they have learned from trusted others.

¹⁵¹ TALI SHAROT, *THE INFLUENTIAL MIND: WHAT THE BRAIN REVEALS ABOUT OUR POWER TO CHANGE OTHERS* 25 (2017).

¹⁵² *Id.*

¹⁵³ CAILIN O’CONNOR & JAMES OWEN WEATHERALL, *THE MISINFORMATION AGE: HOW FALSE BELIEFS SPREAD* 8 (2019).

In the case of judges, studies have demonstrated that this conformity problem exists in the context of the federal sentencing guidelines.¹⁵⁴ Researchers looked at the sentencing data for federal judges throughout the country, surveyed 262 federal district court judges, and interviewed 314 individuals involved in the federal criminal legal system (judges, prosecutors, defense attorneys, etc.).¹⁵⁵ The study found that within judicial communities, defined as the district in which the judge presides, there was considerable conformity for which individual-level predictors (such as the race, age, and/or gender of the defendant) could not account.¹⁵⁶ For example, “downward departures [from the federal sentencing guidelines] were awarded in only 4% of cases in the Northern District of Mississippi, but they were meted out in 31% of the cases in the Districts of Connecticut and Eastern New York.”¹⁵⁷ Furthermore, judges within the same districts were more likely to perceive the federal sentencing guidelines in the same way: as either restrictive or normative.¹⁵⁸ As this study demonstrates, judges are likely to conform their positions to those of their peers, both in terms of belief and practice.

What if, however, we are not certain about the information? Or we trust the wrong people? Or everyone around us is wrong? For years, people were completely certain that the earth was flat, that leeches could cure disease, and that a bitemark could be matched exclusively to a single individual—all of which we now know to be false. Adherence to these beliefs and the subsequent ability to change them is a complex interrelationship among how firmly we hold a belief, how much we trust the person with the new information, and how our community may perceive a change in our belief.¹⁵⁹

False beliefs are “beliefs that are inconsistent with the available evidence, and which are even widely known to be inconsistent with that evidence.”¹⁶⁰ In contrast, a true belief is one that “conform[s] with and [is] supported by the available evidence.”¹⁶¹ Two scenarios can arise with false beliefs: a person (1) genuinely believes the false belief; or (2) does not believe the false belief but because of the conformity problem will adopt the false belief. Both scenarios are relevant to this discussion

¹⁵⁴ Ulmer & Johnson, *supra* note 128, at 255–56.

¹⁵⁵ *Id.* at 266–67.

¹⁵⁶ *See id.* at 277.

¹⁵⁷ *Id.* at 277–78.

¹⁵⁸ *Id.* at 278.

¹⁵⁹ O’CONNOR & WEATHERALL, *supra* note 153, at 25; Ulmer & Johnson, *supra* note 128.

¹⁶⁰ O’CONNOR & WEATHERALL, *supra* note 153, at 7.

¹⁶¹ *Id.* at 43.

because, in either scenario, the conformity problem will be a roadblock to reaching and acting consistent with the true belief.

A few areas of study regarding the impact of social factors on the formulation of true or false beliefs are particularly salient for our understanding of how judges make decisions about forensic science, around which this Section is organized. First is a discussion of how the conformity problem operates, even in the face of evidence that the normative belief is wrong, followed by a discussion of how the strength of belief affects the ability to change. This Section concludes by examining how individuals who wish to perpetuate the false belief can negatively and substantially impact the changing of a norm.

1. The Conformity Problem in the Context of Judicial Decision-Making

The desire to conform, even if that means conforming with something antithetical to one's beliefs, has been long understood to be a part of the human experience. In a classic study on conformity from the 1950s, individuals in groups of eight were given cards that had one vertical line on the left and three vertical lines on the right.¹⁶² The lines on the right were of varying lengths, but one of the lines on that side was the same length as the line on the left.¹⁶³ The individuals were asked which line on the right side was the same length as the line on the left side. In the group of eight "participants," one person was an actual participant in the study and the other seven were confederates whom the researchers instructed to answer the question similarly incorrectly.¹⁶⁴ After hearing seven people give the same wrong answer, the subject, answering last, gave the wrong answer one-third of the time.¹⁶⁵

¹⁶² Solomon E. Asch, *Effects of Group Pressure Upon the Modification and Distortion of Judgments*, in *GROUPS, LEADERSHIP AND MEN: RESEARCH IN HUMAN RELATIONS* 177, 177-79 (Harold Guetzkow ed., 1951).

¹⁶³ *Id.* at 180.

¹⁶⁴ *Id.* at 178.

¹⁶⁵ *Id.* at 181. In a more recent study on conformity, subjects were shown a documentary and then given a test on what they remembered from the documentary. Micah Edelson et al., *Following the Crowd: Brain Substrates of Long-Term Memory Conformity*, 333 *Sci.* 108 (2011). Days after this initial test, subjects were asked to recall the same information, but this time, half were presented with answers from their "peers" (this was a study manipulation, not from actual subjects or peers), all of whom gave incorrect answers. *Id.* Those subjects then gave the wrong answer 68.3% of the time, when they had previously given the correct answer. *Id.* In contrast, the control subjects were not provided with incorrect "peer" answers, and these subjects only gave wrong answers 15.5% of the time. *Id.*

The participants in this study likely did not have a strong emotional tie to answering which line was as long as the other line, but nevertheless, in the face of clear evidence to the contrary, one-third of the people were willing to provide what they knew to be a wrong answer, just so they would be answering the same as the group. People conform because it feels good; having a different belief than everyone around us is uncomfortable,¹⁶⁶ even when they are complete strangers. Imagine how uncomfortable it is when you are saying that your peers have been deciding an issue wrong for decades, such as is the case with much of forensic science.

The following example from nineteenth-century medical research is illustrative of the problems faced in trying to change judicial decision-making about forensic science. In 1846, a Hungarian physician ran an obstetrics clinic for a Viennese hospital when he realized that women who gave birth at his hospital died as a result of complications 10% of the time, while women at a neighboring midwifery clinic died of complications only 3–4% of the time.¹⁶⁷ The doctor soon noticed a connection between his medical staff doing autopsies shortly before assisting with births and the death rate,¹⁶⁸ so he mandated that the medical staff wash their hands before assisting with births.¹⁶⁹ The death rate in this clinic plummeted and the Hungarian doctor published his findings in prominent medical journals.¹⁷⁰ To the doctor's surprise, his ideas about handwashing were rejected by other doctors despite the overwhelming evidence of their successes.¹⁷¹ The doctors, largely upper-class gentlemen, "were offended by the implication that their hands were unclean, and they questioned the scientific basis of his [conclusions], which did not accord with their theories of disease."¹⁷²

This Hungarian doctor's studies demonstrated that failing to wash one's hands before performing a medical procedure could have substantial real-world consequences, including death, but doctors still ignored the information. By embracing the studies, the doctors would be admitting that for decades they had been acting in a manner that jeopardized patient health. And why would an individual doctor adopt this rogue doctor's recommendations—the community of doctors was

¹⁶⁶ O'CONNOR & WEATHERALL, *supra* note 153, at 84.

¹⁶⁷ *Id.* at 77–78.

¹⁶⁸ This was before we knew about bacteria and viruses, and before hospitals had signs about handwashing in every room.

¹⁶⁹ O'CONNOR & WEATHERALL, *supra* note 153, at 78.

¹⁷⁰ *Id.*

¹⁷¹ *Id.* at 79.

¹⁷² *Id.* at 78.

resolutely rejecting the recommendations.¹⁷³ This is not unlike the circumstances surrounding forensic science; many people have been convicted, placed on death row, and possibly executed because of faulty forensic science. Yet judges are still willing to admit the evidence today without critical analysis, even though they should have reason to believe the evidence is unreliable or, at best, has not been shown to be reliable.¹⁷⁴

Over the last fifteen years, judges generally have taken one of two positions regarding FTE: (a) the evidence is not flawed, and the judge will admit it without question; or (b) the evidence is flawed, but the judge will admit it anyway. Judges in the first category, even when presented with evidence regarding the problems with FTE, will ignore the issues and permit the evidence to go before a jury without any limitations.¹⁷⁵ Judges in the second category will admit the evidence with limitations, but the limitations are frequently meaningless to a jury and the testimony has the same impact as it would without the limitation.¹⁷⁶

¹⁷³ *Id.* at 84.

¹⁷⁴ Judge Saris, a judge with a predisposition to want to engage with these issues, see discussion *infra* Section IV.C (discussing her decision in *United States v. Monteiro*), stated at a conference on evidentiary rules, “I am one of those judges who have, and I cringe now, let in bitemark identification, ballistic identification . . .” *Advisory Committee on Evidence Rules: Symposium on Forensic Expert Testimony, Daubert, and Rule 702*, 86 FORDHAM L. REV. 1463, 1535–36 (2018) [hereinafter *Advisory Committee on Evidence Rules*].

¹⁷⁵ *United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1122–23 (D. Nev. 2019); *United States v. Casey*, 928 F. Supp. 2d 397, 400 (D.P.R. 2013).

¹⁷⁶ An example of this is a judge permitting an expert to testify that a casing can be matched to a particular firearm to a “degree of ballistics certainty” rather than “100% certainty.” See, e.g., *United States v. Ashburn*, 88 F. Supp. 3d 239, 249 (E.D.N.Y. 2015) (holding that a qualified FTE expert can testify to a “reasonable degree of ballistics certainty”); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1180 (D.N.M. 2009) (holding that the expert may testify “within a reasonable degree of certainty in the firearms examination field”); *United States v. Monteiro*, 407 F. Supp. 2d 351, 375 (D. Mass. 2006) (holding that if the expert meets the specified qualifications he or she may testify to a degree of “ballistic certainty”). While this may be meaningful to the court, whether the expert says she is “100% certain” or is “testifying to a degree of ballistics certainty,” the jury hears: “The expert is certain that the casing matches the firearm.”

A common pattern emerges in many decisions that critique or limit FTE. The judge often will spend several pages of the decision discussing the following problems with FTE:¹⁷⁷

1. FTE does not have known error rates;
2. FTE is entirely subjective without any objective criteria;
3. The primary peer-reviewed journal, AFTE Journal, does not have blind peer review and the publishers and reviewers have a financial stake in the continued use of FTE evidence in the courts;¹⁷⁸ and
4. No research has demonstrated the foundational premise of FTE: that firearm toolmarks can be individualized.¹⁷⁹

After describing these seemingly insurmountable problems, particularly under *Daubert*, the judge will include a paragraph or two similar to the following:

While these critics of the science underlying ballistic toolmark analysis raise legitimate concerns about whether the process has been demonstrated to be sufficiently reliable to be called a “science,” the defenders of the process—and every federal court to have examined the issue in a written opinion (albeit with considerable differences in the amount of detail in the analysis)—have concluded that it is sufficiently plausible,

¹⁷⁷ See, e.g., *United States v. Simmons*, No. 2:16cr130, 2018 U.S. Dist. LEXIS 18606, at *16–17, *22–24 (E.D. Va. Jan. 12, 2018) (allowing examiner to testify “to a reasonable degree of ballistic or technical certainty” in her conclusions despite the court, the government, and expert acknowledging the subjectivity of FTE); *United States v. McCluskey*, No. 10-2734 JCH, 2013 U.S. Dist. LEXIS 203723, at *26–29, *32–36 (D.N.M. 2013) (concluding that insufficient data exists to determine a definitive error rate and noting the inherent subjectivity of FTE, but allowing examiner to testify “that she has reached her conclusions to ‘a practical certainty,’ or to a ‘practical impossibility’ of dissimilar origin”); *United States v. Diaz*, No. CR 05-00167 WHA, 2007 U.S. Dist. LEXIS 13152, at *15, *23, *41–42 (N.D. Cal. Feb. 12, 2007) (recognizing the issues in testing a technique that relies on subjective standards and inability to calculate a known error rate, but admitting testimony with the limitation that the expert can only make his conclusions to a reasonable degree of certainty in the ballistics field).

¹⁷⁸ The Association of Firearm and Tool Mark Examiners (AFTE) publishes the AFTE Journal. *What is the AFTE Journal?*, ASS’N OF FIREARM & TOOL MARK EXAM’RS, <https://afte.org/afte-journal/what-is-the-journal> (last visited Jan. 20, 2020). Though it claims to be peer reviewed, the AFTE Journal’s peer-review process is not comparable to that of a typical scientific journal. See Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725, 754–56 (2011) (discussing the failure of the AFTE journal to meet the standards usually associated with scientific publication, such as the lack of blind peer review, the selection of peer reviewers from AFTE members on the editorial board, and the limited dissemination of its articles to the wider scientific community).

¹⁷⁹ This is not a comprehensive list of known issues with FTE, but this is discussed in more detail, *infra* Part III.

relevant, and helpful to the jury to be admitted in some form.¹⁸⁰

When presented with the evidence detailing the substantial issues with FTE, those issues are often jettisoned in favor of a position consistent with “every federal court to have examined the issue.” The norm is to admit the evidence.

Though some judges claim they are admitting the evidence because of precedent, precedent was not binding for most of these decisions. Some of the authority upon which the judges rely is persuasive, and for systemic legitimacy, relying on persuasive authority has benefits to the criminal legal system. This argument for systemic legitimacy, however, does not justify the continued reliance on evidence that scientists and other examiners have found to be faulty. In addition, *Daubert* is a case-specific inquiry, designed to permit the exclusion of evidence if the evidence is unreliable in the case at issue, regardless of prior decisions.¹⁸¹ Furthermore, appellate courts review trial court decisions regarding the admission of expert testimony for abuse of discretion.¹⁸² Few appellate courts will overturn a trial court’s decision regarding expert testimony. Therefore, though judges in these decisions frequently rely on precedent, often there is no binding authority requiring the admission of this type of evidence.

Looking at the conformity problem from an economic perspective can be helpful: a person will change their belief when the cost of conforming to that belief becomes greater than the cost of changing that belief.¹⁸³ A judge who has a preference for legal change or who has a strong belief that the state of the law is wrong is more likely to deviate from the norm.¹⁸⁴ In 1970, Judge Skelly Wright deviated from the norm

¹⁸⁰ United States v. Willock, 696 F. Supp. 2d 536, 568 (D. Md. 2010).

¹⁸¹ See discussion of *Daubert*, *supra* Section II.B.2. Courts have noted that evidence previously admitted under *Frye* is still subject to scrutiny under *Daubert*. See, e.g., Coble v. State, 330 S.W.3d 253, 276 n.56 (Tex. Ct. App. 2010) (“[C]ourts do not ‘grandfather in’ expert testimony in a particular field or by a particular witness simply because the court has admitted expert testimony in that field or by that witness in the past.”); United States v. Green, 405 F. Supp. 2d 104, 118 (D. Mass. 2005) (noting that refusing to reexamine evidence previously considered to be generally accepted would be “equivalent to ‘grandfathering old irrationality’” (citing United States v. Hines, 55 F. Supp. 2d 62, 68 n.13 (D. Mass. 1999))); United States v. Saelee, 162 F. Supp. 2d 1097, 1105 (D. Alaska 2001) (“[T]he fact that [expert] evidence has been generally accepted in the past by courts does not mean that it should be generally accepted now, after *Daubert* and *Kumho Tire*.”).

¹⁸² Gen. Elec. Co. v. Joiner, 522 U.S. 136, 143 (1997) (holding that the admissibility of expert testimony “is reviewable under the abuse-of-discretion standard”).

¹⁸³ Harnay & Marciano, *supra* note 128, at 407–09.

¹⁸⁴ *Id.* at 407; Thomas J. Miceli & Metin M. Cosgel, *Reputation and Judicial Decision-Making*, 23 J. ECON. BEHAV. & ORG. 31, 33–35 (1994).

and found a warranty of habitability in rental leases—a warranty that had never previously existed.¹⁸⁵ In so deciding, Judge Skelly Wright stated,

I didn't like what I saw, and I did what I could to ameliorate, if not eliminate, the injustice involved in the way many of the poor were required to live in the nation's capital. I offer no apology for not following more closely the legal precedent which had cooperated in creating the conditions that I found unjust.¹⁸⁶

The combination of how other judges will perceive a deviation from precedent, along with their personal view of how the case should be decided and the extra work necessary to deviate from precedent, will often dictate whether a judge will conform or deviate.¹⁸⁷

Thus, deviating is costly not only because it forces one to stand apart but also because it often takes more time and energy. Judge Rakoff recently described the pressures of presiding over *Daubert* hearings:

The problem with *Daubert* and [Federal Rules of Evidence] Rule 702 is that these hearings, if they're done right, take a lot of time. And judges are very busy and they always have more on their docket than they can handle in an ideal fashion. I had a two-week *Daubert* hearing in a case. It was fascinating. I loved every minute of it, but who can afford two weeks on a single *Daubert* hearing?¹⁸⁸

To be clear, Judge Rakoff is predisposed to want to engage in evidentiary issues—he has served on committees addressing these issues and written about them.¹⁸⁹ Even for him, though, paying close attention to these issues has a real cost. Relying on the norm and just admitting the

¹⁸⁵ *Javins v. First Nat'l Realty Corp.*, 428 F.2d 1071, 1072–73 (D.C. Cir. 1970).

¹⁸⁶ Edward H. Rabin, *The Revolution in Residential Landlord-Tenant Law: Causes and Consequences*, 69 CORNELL L. REV. 517, 549 (1984).

¹⁸⁷ Harnay & Marciano, *supra* note 128, at 407; Miceli & Cosgel, *supra* note 184, at 49.

¹⁸⁸ *Advisory Committee on Evidence Rules*, *supra* note 174, at 1537–38.

¹⁸⁹ *Jed Rakoff*, COLUM. LAW SCH., <https://www.law.columbia.edu/faculty/jed-rakoff> (last visited Feb. 11, 2020). Many of the judges who have issued rulings limiting forensic science expert testimony are judges who have a specialized interest in complex evidentiary issues and have served on committees, lectured, or written in the area. In addition to Judge Rakoff, Judge Saris served on the PCAST advisory committee; Judge Gertner has written law review articles in the area; and Judge Grimm has served on committees and has spoken at the Ninth Circuit Judicial Conference urging trial judges to take more time to assess the admissibility of forensic science evidence. Maria Dinzeo, *Skepticism of Forensic Methods Urged at 9th Circuit Conference*, COURTHOUSE NEWS SERV. (July 18, 2017), <https://www.courthousenews.com/skepticism-forensic-methods-urged-9th-circuit-conference/>. These are judges who are particularly interested in the area, willing to hear both sides of this issue, and more likely to change their beliefs.

evidence almost always has a lower cost to the judge in terms of time, energy, and conformity.

Furthermore, judges may not have sufficient motivation to take on the costs associated with changing the way they approach forensic science. Studies show that personal incentives can substantially impact whether a person is willing to deviate.¹⁹⁰ For the vast majority of judges, the decision to admit forensic science evidence is not personal, just as the decision whether to wash one's hands was not personal to the Hungarian doctors. Without a personal hook in the decision, the cost of conforming remains low.

2. Forensic Science as a Strongly Held Belief

Critical to this cost-benefit determination are (1) a person's confidence in that old (false) belief, and (2) a person's confidence in the new (true) evidence.¹⁹¹ Some beliefs are so ingrained in who we are as individuals that the new evidence and our confidence in that evidence have to be particularly strong to move us. For example, a person's religious affiliation is often a very strongly held belief. Though one might not initially think the accuracy of forensic science would be a strongly held belief, it must be viewed in the context of the cases in which it is presented: "scientific" evidence presented by prosecutors and law enforcement to support allegations of often serious and violent offenses.¹⁹² Many judges do have strong beliefs in that context, particularly when judges have relied on forensic science for decades.

For example, in Santa Clara County, California, a recent study examined five years of criminal jury trial appeals and found "a pattern of judicial conduct that favored prosecutors, with incidents occurring at nearly every step of the proceedings."¹⁹³ The Mercury News, a local newspaper, reviewed 727 cases, and in more than fifty of those cases, "judges allowed prosecutors to introduce questionable—and often improper—evidence."¹⁹⁴ In all, the paper found "more than 100

¹⁹⁰ O'CONNOR & WEATHERALL, *supra* note 153, at 89–90.

¹⁹¹ SHAROT, *supra* note 151.

¹⁹² Political pressure on judges, even those who are not elected to office, strongly favors prosecutors and disfavors criminal defendants. "In other words, to the extent that a trial judge facing an admissibility determination is influenced by political pressure, it is likely to push her toward admitting the evidence." Stephanie L. Damon-Moore, Note, *Trial Judges and the Forensic Science Problem*, 92 N.Y.U. L. REV. 1532, 1566 (2017).

¹⁹³ Fredric N. Tulskey, *Part Four: How Judges Favor the Prosecution*, MERCURY NEWS (Jan. 31, 2007, 10:36 AM), <https://www.mercurynews.com/2007/01/31/part-four-how-judges-favor-the-prosecution/>.

¹⁹⁴ *Id.*

instances when the appellate courts found that the trial judges erred in ways that helped prosecutors, and more than [forty] additional instances of troubling conduct that the appellate courts declined to assess.”¹⁹⁵

In addition, though most criminal cases have a defense attorney on one side and a prosecutor on the other, the judges very clearly tilt in one direction: on the federal bench, former prosecutors outnumber former defense attorneys four to one.¹⁹⁶ At least in the federal courts, the chances are far higher that a forensic science issue will be before a judge who previously has relied on this evidence rather than before a judge who has challenged this evidence.

These issues are exacerbated when one considers the likely impact of “belief perseverance,” a form of bias in which individuals hold onto initial beliefs even in the face of evidence that entirely contradicts that earlier belief.¹⁹⁷ Furthermore, the longer a person holds a belief, the more difficult it is to change.¹⁹⁸ In this context, judges who have believed in the reliability of forensic science for years and years, and particularly those judges who are former prosecutors, are likely to have a strongly held belief that forensic science is reliable and foundationally sound.

Finally, once conformity shapes a person’s belief, the memory and brain can change such that even after the person is told that the basis of the original belief was false, the belief nevertheless remains unchanged. A study of the long-term impacts of conformity on the brain demonstrated that social influence perpetuating a false belief resulted in changes to the study subjects’ brains such that the subjects were resistant to reversing the impacts of the social influence even after they knew of its falsity.¹⁹⁹

This combination of familiarity with forensic science from a prosecutorial perspective, along with a history of belief in the reliability of forensic science, results in a judiciary that, in general, strongly believes that this evidence is reliable. The difficulty of moving someone off a strongly held belief is a substantial contributor to the conformity problem.

¹⁹⁵ *Id.*

¹⁹⁶ Neily, *supra* note 150.

¹⁹⁷ See Huang Shiyuan, *Cognitive Biases That Lead to Wrongful Convictions: Illustrated by Twenty-Three Erroneous Chinese Cases*, 54 CAL. W. L. REV. 103, 108 (2018); Moriarty, *supra* note 127, at 320.

¹⁹⁸ Shiyuan, *supra* note 197, at 108.

¹⁹⁹ Edelson et al., *supra* note 165, at 110.

3. The Role of Prosecutors in Maintaining the Conformity Problem

Finally, a contributing factor to this conformity problem is the influence of prosecutors who seek to admit forensic science evidence in criminal trials. The issue this Article seeks to address would not be a problem but for the prosecutors seeking to admit this evidence at trial and, significantly, asking that the prosecution experts be permitted to testify to unsupported degrees of certainty. This directly contributes to the conformity problem because when prosecutors seek to perpetuate this false belief that forensic sciences are reliable, it gives judges a reason to hold onto that false belief and continue on the path of least resistance: maintain the norm and admit the evidence.²⁰⁰

In the adversarial system that is the bedrock of the criminal legal system, each side is given an equal opportunity to advance and defend its position on issues.²⁰¹ In addition, not only in the adversarial system but also generally, people prefer to give everyone's opinion equal weight, even when the expertise is not equal on both sides.²⁰² Therefore, a prosecutor advancing a position that is not based on scientific evidence will often receive equal weight when compared with a defense attorney armed with the PCAST Report, NAS Report, and any other applicable scientific articles or reports. This is problematic because

²⁰⁰ The prosecutor's influence may be further compounded based on her role in the criminal legal system and connection with the court. The judge may come from the same office as the prosecutor now presenting such evidence for admission, or even if not from the same office, the judge may have a lot of confidence in that prosecutor who is urging her to follow all the other judges and admit the evidence. Compare *United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1122–23 (D. Nev. 2019) (admitting testimony regarding FTE without any limitations), with *United States v. Willock*, 696 F. Supp. 2d 536, 574 (D. Md. 2010) (preventing the firearms toolmark examiner from testifying to any degree of certainty). The judge in *Romero-Lobato* served as a prosecutor for a decade before entering private practice and then becoming a judge. Though the magistrate judge who presided over the *Daubert* hearing in *Willock* was a prosecutor for four years a few decades before taking the bench, he has developed a specialty in complex discovery and evidentiary issues since becoming a judge. See *Judge Larry R. Hicks*, U.S. DIST. CT., DIST. OF NEVADA, <https://www.nvd.uscourts.gov/court-information/judges/judge-larry-r-hicks/> (last visited Jan. 7, 2020); *Paul W. Grimm, District Judge*, U.S. DIST. CT., DIST. OF MD., <https://www.mdd.uscourts.gov/paul-w-grimm-district-judge> (last visited Jan. 7, 2020); *SDTX Bench Bar Conference Bio—Judge Paul W. Grimm*, U.S. DIST. & BANKR. CT., S. DIST. OF TEX., <https://www.txs.uscourts.gov/page/sdtx-bench-bar-conference-bio-judge-paul-w-grimm> (last visited Jan. 10, 2020).

²⁰¹ Merely the fact that both sides are given equal time can be problematic when one side has evidence and science on its side while the other does not; it creates a false impression that both sides are equally supported. See O'CONNOR & WEATHERALL, *supra* note 153, at 158–60 (explaining how the “Fairness Doctrine” can distort public opinion on issues when two sides are treated equally despite a lack of equality in the evidentiary support for the opinions).

²⁰² SHAROT, *supra* note 151, at 190.

though both sides are often given equal weight, the evidence supports just one side.

This scenario, however, envisions this litigation happening under a best-case scenario: prosecutors and defense attorneys equally matched and equally armed with the necessary information. Unfortunately, much of the litigation in criminal cases is not happening in this idealized courtroom. In many cases, the defense attorney may not even be aware of the problems with forensic science or have the resources to contest the evidence.²⁰³ But even if the defense attorney is well-equipped to address forensic science issues, prosecutors still have an obligation to only seek admission of reliable forensic science evidence.²⁰⁴

²⁰³ Several law review articles have identified that many public defender offices and individual defense attorneys do not have the resources to contest expert testimony on forensic science issues. See, e.g., Jessica D. Gabel, *Realizing Reliability in Forensic Science from the Ground Up*, 104 J. CRIM. L. & CRIMINOLOGY 283, 336–37 (2014) (discussing the disparity in access to and funding of defense experts); Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1, 89–90 (2009) (discussing defense attorneys' lack of knowledge and expertise in forensic science and the trial courts' failure to appoint defense experts); Paul C. Giannelli, *Ake v. Oklahoma: The Right to Expert Assistance in a Post-Daubert, Post-DNA World*, 89 CORNELL L. REV. 1305, 1315–16 (2004) (discussing the need to appoint defense experts in all criminal trials because “[f]ew defense attorneys can deal with this type of sophisticated evidence . . . without expert assistance,” and the disparity between prosecution and defense access to and funding for experts); Mark Loudon-Brown, *Garbage In, Garbage Out: Revising Strickland as Applied to Forensic Science Evidence*, 34 GA. ST. U. L. REV. 893, 894 (2018) (noting that a large caseload combined with a lack of familiarity with scientific issues often results in defense attorneys avoiding forensic science issues); Peter J. Neufeld & Neville Colman, *When Science Takes the Witness Stand*, 262 SCI. AM. 46, 52–53 (1990) (discussing the lack of funding and time that defense attorneys have to address complex scientific evidence and the failure of trial judges to authorize funds for defense experts).

²⁰⁴ All lawyers have ethical obligations to avoid making false statements of material fact or law to the court or a third person. MODEL RULES OF PROF'L CONDUCT r. 3.3, 4.1 (AM. BAR ASS'N 2019). Prosecutors are ministers of justice, not simply advocates and have additional special responsibilities and obligations to “see that the defendant is accorded procedural justice, that guilt is decided upon the basis of sufficient evidence, and that special precautions are taken to prevent and to rectify the conviction of innocent person.” *Id.* r. 3.8 cmt. Prosecutors must “make timely disclosure to the defense of all evidence or information known to the prosecutor that tends to negate the guilt of the accused or mitigates the offense” and promptly disclose credible and material evidence creating a reasonable likelihood a convicted defendant did not commit the offense of which they were convicted. *Id.* r. 3.8; see also, Lara Bazelon, *Ending Innocence Denying*, 47 HOFSTRA L. REV. 393 (2018) (discussing the ethical obligation of prosecutors to act as ministers of justice and pursue exoneration when a defendant is convicted based on faulty forensic science evidence or proven innocent by conclusive forensic science evidence); Jane Campbell Moriarty, “Misconvictions,” *Science, and the Ministers of Justice*, 86 NEB. L. REV. 1, 20–30 (2007) (criticizing prosecutorial use of unreliable forensic science evidence and expert witnesses and advocating for a higher ethical standard requiring a prosecutor to fulfill an additional gatekeeping role and avoid the use of such unreliable evidence).

Cases involving forensic science are often serious violent crimes, and prosecutors' desire to convict people for those crimes is understandable. To admit or use forensic science, however, without first establishing its reliability can result in wrongful convictions and the factually guilty party remaining free to commit additional violent offenses. For example, in the Robert Lee Stinson case, discussed *supra*, the newly discovered DNA evidence pointed toward another individual, Moses Price, Jr.²⁰⁵ Mr. Price ultimately confessed to the 1985 crime for which Mr. Stinson had been wrongfully convicted.²⁰⁶ At the time that Mr. Price was identified as the true perpetrator, he was serving a thirty-five-year sentence for a 1991 murder—a murder committed after the offense for which Mr. Stinson was wrongly convicted.²⁰⁷ Using faulty forensic science to convict the wrong person can have serious ripple effects far beyond injustice of a the wrongfully convicted individual.

Since the NAS Report and PCAST Report were first released, leading federal prosecutors have resisted their conclusions. In response to the PCAST Report, former U.S. Attorney General Loretta Lynch (serving under President Barack Obama) stated:

[The Department of Justice] remain[s] confident that, when used properly, forensic science evidence helps juries identify the guilty and clear the innocent, and the Department believes that the current legal standards regarding the admissibility of forensic evidence are based on sound science and sound legal reasoning. . . . While we appreciate [the PCAST Report's] contribution to the field of scientific inquiry, the department will not be adopting the recommendations related to the admissibility of forensic science evidence.²⁰⁸

Former U.S. Attorney General Jeff Sessions (serving under President Donald Trump), who disbanded the National Commission on Forensic Science in 2017, stated at the Senate hearing regarding the NAS Report, "I don't think we should suggest that those proven scientific principles that we've been using for decades are somehow uncertain

²⁰⁵ Mike Johnson, *Man Charged in 1984 North Side Slaying*, MILWAUKEE WIS. J. SENTINEL (Apr. 3, 2012), <https://archive.jsonline.com/news/crime/man-charged-in-1984-north-side-slaying-8r54lrl-148613205.html>.

²⁰⁶ *Id.*

²⁰⁷ *Id.*

²⁰⁸ Adam B. Schniderman, *Prosecutors Respond to Calls for Forensic Science Reform: More Sharks in Dirty Water*, 126 YALE L.J.F. 348, 349 (2016), https://www.yalelawjournal.org/forum/prosecutors-respond-to-calls-for-forensic-science-reform#_ftnref6 (quoting Kira Lerner, *Attorney General to Ignore New Report Finding That Commonly-Used Forensics are Bogus*, THINK PROGRESS (Sept. 21, 2016)).

....”²⁰⁹ Since disbanding the Commission, DOJ officials have continued to make comments providing blanket support for forensic science and attempted to discredit any challenge to such evidence. Then-Deputy Attorney General Rod Rosenstein addressed the attendees at the National Symposium on Forensic Science in 2018:²¹⁰

Most of you work on the front lines of the criminal justice system, where forensic science has been under attack in recent years. Some critics would like to see forensic evidence excluded from state and federal courtrooms. . . . Critics argue that the methods have not undergone the right type or amount of validation, or that they involve too much human interpretation and judgment to be accepted as ‘scientific’ methods. Those arguments are based on the false premise that a scientific method must be instrument-based, automated, and quantitative, excluding human interpretation and judgment. . . . The effort stems from an erroneously narrow view of the nature of science and its application to forensic evidence.²¹¹

Not only does Rosenstein misstate the concerns of the NAS Report, PCAST Report, the disbanded Commission, and other critics, he affirms the reliability of forensic science evidence in criminal cases in the face of overwhelming evidence to the contrary.²¹²

Particularly in the case of the DOJ officials, it simply cannot be said that the prosecutors are ignorant of the problems with forensic science and the gross overstatements of reliability. At the same time that then-Deputy Attorney General Rosenstein was lauding forensic science, the DOJ was prohibiting FTE examiners from testifying that they can match ballistics to a single firearm to the exclusion of all other firearms.²¹³ In

²⁰⁹ *Strengthening Forensic Science in the United States: Hearing before the S. Comm. on the Judiciary*, 111th Cong. (2009) (statement of Sen. Jeff Sessions).

²¹⁰ This was an event for local, state, and federal prosecutors co-sponsored by the DOJ, National Association of Attorneys General, and the National District Attorneys Association. *Deputy Attorney General Rod J. Rosenstein Delivers Remarks at the National Symposium on Forensic Science*, U.S. DEPT. OF JUST. (Aug. 7, 2018), <https://www.justice.gov/opa/speech/deputy-attorney-general-rod-j-rosenstein-delivers-remarks-national-symposium-forensic> [hereinafter *Deputy Atty Gen. Rosenstein*].

²¹¹ *Id.*

²¹² Much can be said, and has been said, about prosecutors’ responsibility to do justice and take a different, more just position when it comes to the admission of forensic science in criminal cases. See *supra* note 204 for a discussion of prosecutors’ ethical responsibilities. For purposes of this paper, it is the influence of these prosecutors and their stance that is directly relevant.

²¹³ *Uniform Language for Testimony and Reports for the Forensic Firearms/Toolmark Discipline—Pattern Match Examination*, U.S. DEPT OF JUST. (July 24, 2018),

addition, the FBI is currently undertaking the very studies that the NAS Report, PCAST Report, and other critics have demanded.²¹⁴ The problem is that the DOJ wants this evidence to continue to be admissible while it awaits the results from the studies, even though it has no idea what the error rates are or how reliable the testing is.

Furthermore, the DOJ is providing information to stakeholders, including judges, in ways that can be misleading. For example, at the 2017 Symposium on Forensic Expert Testimony, *Daubert*, and Rule 702, a conference for judges,²¹⁵ Ted Hunt, Senior Advisor on Forensic Science at the DOJ, claimed that the latent print study from the American Association for the Advancement of Science (AAAS) rejected the issues of foundational validity that the PCAST Report raised.²¹⁶ Furthermore, he stated, “The PCAST approach . . . is not consistent with the scientific mainstream’s holistic approach to method validation that is embodied in the currently governing international standard.”²¹⁷ These statements gave the impression that the PCAST Report’s conclusions are in direct conflict with the mainstream scientific community when, in fact, the AAAS report largely agreed with the PCAST Report—the title of the AAAS news release was: “Fingerprint Source Identity Lacks Scientific Basis for Legal Certainty.”²¹⁸ Similarly, the PCAST Report concluded that, though the FBI is now undertaking appropriate studies of latent fingerprint examination, current error rates and foundational validity have not been established.²¹⁹

Creating confusion about the debate and asserting that both sides of the debate have evidentiary support increases the likelihood the judges will continue to rely on false beliefs; if a judge believes in the efficacy of forensic science evidence and the evidence to the contrary is uncertain, the judge is going to be even less inclined to change her position.²²⁰ This is a substantial contributor to the conformity problem because, as discussed above in Section II.A., a key consideration for whether a judge will deviate or conform is the relative economic cost. If

<https://www.justice.gov/olp/page/file/1083671/download>; *Deputy Att’y Gen. Rosenstein*, *supra* note 210.

²¹⁴ See *Advisory Committee on Evidence Rules*, *supra* note 174, at 1484–85.

²¹⁵ The Judicial Conference Advisory Committee of Evidence rules sponsored this Symposium. *Id.* at 1463.

²¹⁶ *Id.* at 1520.

²¹⁷ *Id.* at 1521.

²¹⁸ Anne Q. Hoy, *Fingerprint Source Identity Lacks Scientific Basis for Legal Certainty*, AM. ASS’N FOR THE ADVANCEMENT OF SCI. (Sept. 14, 2017), <https://www.aaas.org/news/fingerprint-source-identity-lacks-scientific-basis-legal-certainty>.

²¹⁹ PCAST REPORT, *supra* note 7, at 17–19.

²²⁰ See O’CONNOR & WEATHERALL, *supra* note 153, at 158–60.

prosecutors make it seem as though the reliability of forensic science is a contested issue, the economic cost of maintaining the norm, the position for which the prosecutor is advocating, becomes lower. If the prosecutor admits that the forensic expert should be limited in their testimony, then the economic cost of deviating is lower.²²¹

One could compare the actions of those prosecutors who know that forensic science is flawed and yet still seek its admission with the actions of cigarette companies, which have attempted to obfuscate the connection between smoking and cancer since the 1950s. Just as smoking is a public health issue that can result in serious health consequences, flawed forensic science can result in wrongful convictions and substantial losses of liberty. After a *Reader's Digest* article in 1952 demonstrated a link between cigarette smoking and lung cancer (and there was a massive sell-off of stock in tobacco companies), the tobacco companies created the "Tobacco Strategy."²²² "The goal was . . . to create the appearance of uncertainty: to find, fund, and promote research that muddled the waters, made the existing evidence seem less definitive, and gave policy makers and tobacco users just enough cover to ignore the scientific consensus."²²³

This is substantially like what some prosecutors are doing around forensic science. Sowing a climate of confusion and uncertainty regarding the reliability of this evidence is very similar to what tobacco companies attempted to do with the link between cancer and smoking—if people are unsure of the connection, they will feel comfortable relying upon the previously held belief and are unlikely to change course.²²⁴ Furthermore, because some prosecutors continue to perpetuate this idea that there is a debate as to the efficacy of these areas of forensic science, both "sides" to the debate receive equal attention,²²⁵ and judges will continue to view the choices as having equal bases in evidence.

Thus, the conformity problem persists where judges feel pressure to go along with their peers, they believe strongly in the reliability of forensic science, and prosecutors regularly attempt to create a false sense of equality in the debate over these issues. For most judges, the cost of deviating in these circumstances is high. Some judges, however, are not so inclined. Finding those judges, convincing those judges, and

²²¹ Some forensic science evidence, such as bitemark analysis, should not be introduced at all. If prosecutors cease asking judges to admit this evidence, then judges will not have to grapple with whether to admit the evidence.

²²² O'CONNOR & WEATHERALL, *supra* note 153, at 93–95.

²²³ *Id.* at 95.

²²⁴ *See id.* at 101–02.

²²⁵ *Id.* at 157–59.

getting favorable rulings over and over and over again is how the conformity problem can start working in favor of the true belief, rather than the false belief.

IV. INSTANCES OF THE CONFORMITY PROBLEM IN JUDICIAL DECISION-MAKING

The conformity problem may be pushing judges to admit FTE evidence even when the judge has concerns about doing so. The thinking is, “If every other judge is admitting it, why shouldn’t I?” If enough judges start to deviate, though, and substantially limit or exclude the evidence, the norm can change, and the conformity problem is no longer a problem.²²⁶ This Part demonstrates that once courts—even a small number of them—start deviating from the norm, other courts will follow.

In mid-1600s England, smallpox was prevalent, and for twenty to sixty percent of the people who contracted it, it was fatal.²²⁷ One woman in the British aristocracy, Lady Mary Wortley Montagu, had contracted smallpox in her twenties. Even though she survived, she was horribly scarred and lost her brother to the disease.²²⁸ Soon after, Lady Mary traveled to Turkey where she learned of smallpox variolation, which is not dissimilar from today’s vaccines. Variolation caused sickness and possible death in a small percentage of people, but for the vast majority, it would make them immune to smallpox.²²⁹ Lady Mary was thrilled with this discovery and tried to bring variolation back to England.²³⁰ Once back in England, however, no one wanted to listen to Lady Mary about variolation; not only was this a foreign procedure but a *woman*

²²⁶ In a study on how pressure to conform impacts memory, the authors created several conditions: one in which all the peers incorrectly disagreed with the subject’s answer; one in which some but not all of the peers incorrectly disagreed with the subject’s answer; and one in which no information was provided about peer answers. Micah Edelson et al., *Brain Substrates of Recovery from Misleading Influence*, 34 J. NEUROSCIENCE 7744, 7745 (2014) (another group in the study was provided random, computer generated answers, but this condition is not relevant to the instant discussion). The subjects who were told that their peers unanimously disagreed with their answer changed their answer to the incorrect answer 68.3% of the time. *Id.* at 7747–48. Those who were told nothing about their peers changed their answer to the incorrect answer 15.5% of the time. *Id.* at 7748. Those who were told some peers had the incorrect answer and some had the same answer provided the incorrect answer 11.8% of the time. *Id.* This study demonstrates that when just some of the peer group agrees with the subject, even if others disagree, the subject will maintain the correct belief rather than conforming with those who disagree.

²²⁷ O’CONNOR & WEATHERALL, *supra* note 153, at 139.

²²⁸ *Id.*

²²⁹ *Id.* at 139–40.

²³⁰ *Id.* at 140.

was advocating for it.²³¹ But Lady Mary found an ally in the Princess of Wales, Princess Caroline, who was hugely popular at the time.²³² Princess Caroline publicly chose to have variolation performed on her children and soon the aristocracy followed, then the doctors, then the public.²³³

Just as one well-respected individual was able to change the way individuals perceived variolation in the 1600s, one well-respected court can change the way judges and other individuals perceive legal and evidentiary issues. This Part starts by examining how *State v. Henderson*,²³⁴ a New Jersey Supreme Court case on eyewitness identification, impacted how eyewitness identification issues are perceived and litigated across the United States. *Henderson* is an exemplar of how a change in one jurisdiction can change norms and the conformity problem beyond the court's jurisdiction. Then, this Part looks at a series of cases from the District of Columbia courts on FTE evidence and how the decisions in that jurisdiction moved from an acceptance of FTE to evidence-based skepticism. Finally, two FTE cases from a Massachusetts federal court demonstrate how deviations in one jurisdiction can have a national impact, even if the local courts do not follow.

A. New Jersey Identification Cases

The *Henderson* decision created a new framework for courts to evaluate whether an eyewitness identification is admissible at trial,²³⁵ establishing a higher standard of review than what the U.S. Supreme Court devised in *Manson v. Brathwaite*.²³⁶ Since the New Jersey Supreme

²³¹ *Id.*

²³² *Id.*

²³³ O'CONNOR & WEATHERALL, *supra* note 153, at 140–41.

²³⁴ *State v. Henderson*, 27 A.3d 872 (N.J. 2011).

²³⁵ The *Henderson* Court rejected the *Manson v. Brathwaite* test, *see discussion infra* note 236, and instead held that for a defendant to get a hearing on the admissibility of an identification, the defendant must merely establish some evidence of suggestiveness, which must be tied to a system variable (such as the actions of the state actors). *Henderson*, 27 A.3d at 288–89. Once a defendant has met this threshold, the burden is on the prosecutor to “offer proof to show that the proffered eyewitness identification is reliable—accounting for system and estimator variables,” which includes factors impacting the witness’s ability to view the suspect and memory. *Id.* at 289. Finally, the burden remains on the defendant to prove a very substantial likelihood of irreparable misidentification. *Id.* In addition to altering the standard, the *Henderson* Court mandated the creation of a jury instruction that incorporated the social science on identifications, which a judge must read in any trial where an identification is at issue. *Id.* at 298–99.

²³⁶ *Manson v. Brathwaite*, 432 U.S. 98 (1977). In *Manson*, the Court established that “reliability is the linchpin in determining the admissibility of identification testimony”

Court issued the *Henderson* decision in 2011, at least 149 state court decisions from outside New Jersey have cited to *Henderson*.²³⁷ In addition, 171 law review articles have cited the decision.²³⁸ *Henderson's* impact has extended far beyond the borders of New Jersey.

The *Henderson* decision and eyewitness identification case law are particularly instructive for the issues relating to admission of forensic science evidence: people, including judges, have believed both eyewitness identifications and forensic science to be reliable, yet they are the most frequent causes of wrongful convictions. Though issues with eyewitness identification have been noted since at least 1932 when a Yale University law professor, Edwin Borchard, accounted for almost seventy instances of innocent individuals being convicted based on erroneous eyewitness identification,²³⁹ the uncovering of wrongful convictions spurred by the advent of DNA evidence brought to light the seriousness of the problem. As the *Henderson* Court noted, "Nationwide, more than seventy-five percent of convictions overturned due to DNA evidence involved eyewitness misidentification.... [I]t has been estimated that approximately 7,500 of every 1.5 million annual convictions for serious offenses may be based on misidentifications."²⁴⁰ In addition, social science research on the unreliability of eyewitness identification has been mounting.²⁴¹

where the defendant alleges that the identification procedure was impermissibly suggestive. *Id.* at 114. In determining whether the identification was reliable, the defendant must first demonstrate that the identification was impermissibly suggestive, and then a court should consider the totality of the circumstances, including opportunity to view the suspect, degree of attention, accuracy of any prior description, level of certainty at the time of the identification, and the time lapse between the incident and the confrontation. *Id.* Unless the circumstances demonstrate "a very substantial likelihood of irreparable misidentification," the evidence will be admissible. *Id.* at 116 (internal citation and quotations omitted).

²³⁷ See *Citing References*—State v. Henderson, WESTLAW EDGE, <https://1.next.westlaw.com> (search for State v. Henderson, 27 A.3d 872 (N.J. 2011); select Citing References tab; and select cases as the content type) (last visited Nov. 13, 2020). As of November 13, 2020, New Jersey courts had cited to *Henderson* in 357 judicial decisions. See *id.* These numbers do not include trial court orders that are not reduced to writing or that are written but not published to Westlaw.

²³⁸ See *Citing References*—State v. Henderson, WESTLAW EDGE, <https://1.next.westlaw.com> (search for State v. Henderson, 27 A.3d 872 (N.J. 2011); select Citing References tab; and select secondary sources as the content type) (last visited Feb. 13, 2020).

²³⁹ NAT'L RSS. COUNCIL, IDENTIFYING THE CULPRIT: ASSESSING EYEWITNESS IDENTIFICATION 9 (2014).

²⁴⁰ State v. Henderson, 27 A.3d 872, 886 (N.J. 2011) (internal quotation marks and citations omitted).

²⁴¹ *Id.* at 886–88. See, e.g., Bruce W. Behrman & Sherrie L. Davey, *Eyewitness Identification in Actual Criminal Cases: An Archival Analysis*, 25 L. & HUM. BEHAV. 475 (2001) (studying police cases to find that suspect identification rates decrease over

Yet, just as forensic science is often persuasive, a witness testifying in court, under oath, that the person sitting at the defense table committed the alleged offense can have a tremendous impact on a jury.²⁴² The conflict between what seems to be reliable based on common sense and what science shows us to be reliable is at the heart of the slow progress to limit both forensic science testimony and eyewitness identification testimony.

The *Henderson* decision did not come to be overnight. Not only did the case itself take eight years from the date of the offense to the date of the seminal decision but it also built on years of decisions in which the New Jersey Supreme Court incrementally incorporated social science into its evaluation of the reliability of identifications. A 1999 case mandated that the trial courts read a jury instruction on cross-racial bias in eyewitness identification in applicable cases.²⁴³ In 2006, the high court required police officers to record identification procedures and

time, are significantly greater when the suspect is the same race as the witness, and are significantly higher for show ups than lineups); Robert K. Bothwell, John C. Brigham & Roy S. Malpass, *Cross-Racial Identification*, 15 PERSONALITY & SOC. PSYCHOL. BULL. 19 (1989) (finding that memory for faces of the same race as the witness is superior to memory for faces of a different race); Amy Bradfield Douglass & Nancy Steblay, *Memory Distortion in Eyewitnesses: A Meta-Analysis of the Post-Identification Feedback Effect*, 20 APPLIED COGNITIVE PSYCH. 859, 864-65 (2006) (finding that confirming feedback after identification caused a witness to inflate their reports of confidence in their identification, memory abilities, and conditions at the time of witnessing an event); Lynn Garrioch & C.A. Elizabeth Brimacombe, *Lineup Administrators' Expectations: Their Impact on Eyewitness Confidence*, 25 L. & HUM. BEHAV. 299 (2001) (finding that eyewitness' confidence changes as a function of the lineup administrator's belief about the perpetrator's identity and that jurors considered identifications equally credible regardless of a witness' level of confidence); Elizabeth F. Loftus, *Unconscious Transference in Eyewitness Identification*, 2 L. & PSYCH. REV. 93 (1976) (finding that eyewitnesses to crimes can mistakenly identify a person seen in another context as the perpetrator of the crime); Elizabeth F. Loftus, *Leading Questions and the Eyewitness Report*, 7 COGNITIVE PSYCH. 560 (1975) (finding that questions asked immediately after observing an event can introduce new information that alters the memory of the event); Roy S. Malpass & Patricia G. Devine, *Eyewitness Identification: Lineup Instructions and the Absence of the Offender*, 66 J. APPLIED PSYCH. 482 (1981) (finding that the confidence of an eyewitness is unrelated to the accuracy of their identification and that biased instructions increase the rate of false identifications); Steven D. Penrod & Brian L. Cutler, *Eyewitness Expert Testimony and Jury Decisionmaking*, 52 L. & CONTEMP. PROBS 43 (1989) (finding that jurors will judge the accuracy of an identification by the confidence of the eyewitness, and the presentation of expert testimony will cause a jury to more carefully consider the identification).

²⁴² See, e.g., Ralph D. Gants & Erik N. Doughty, *Where Science Conflicts with Common Sense: Eyewitness Identification Reform in Massachusetts*, 79 ALB. L. REV. 1617, 1617 (2015).

²⁴³ *State v. Cromedy*, 727 A.2d 457, 467 (N.J. 1999). DNA evidence later exonerated Mr. Cromedy. See also C.J. Stuart Rabner, *Evaluating Eyewitness Identification Evidence in the 21st Century*, 87 N.Y.U. L. REV. 1249, 1256 (2012).

where recordation is infeasible, to prepare a detailed written record.²⁴⁴ A 2007 case mandated a jury instruction “that a witness’s level of confidence, standing alone, may not be an indication of the reliability of the identification.”²⁴⁵ These decisions, slowly building upon and incorporating relevant social science, culminated in *Henderson*’s overhaul of how eyewitness identification evidence was handled.

The *Henderson* case stemmed from a 2003 murder in which a witness made an identification from a photo array.²⁴⁶ The identification was ultimately admitted at trial, and a jury convicted the defendant.²⁴⁷ After a defense appeal, the case was remanded in 2008 for further hearings on the identification, but the New Jersey Supreme Court accepted certification that year, and, after a hearing, requested a special hearing on the social science underlying eyewitness identifications.²⁴⁸ During the ten-day special hearing, the presiding special master heard from multiple eyewitness identification experts and issued a comprehensive report finding that eyewitnesses can make misidentifications under a multitude of circumstances.²⁴⁹ These findings were incorporated into the *Henderson* decision.²⁵⁰ In response to the social science, the Court completely changed how identification procedures are conducted in New Jersey and evaluated during court challenges.²⁵¹ These significant changes came after eight years of litigation and record-building in this case and after decades of litigation in New Jersey.

This landmark decision resulted in almost immediate changes in two states: both the Oregon and Idaho high courts issued decisions mirroring *Henderson* within the next two years.²⁵² In 2015,

²⁴⁴ *State v. Delgado*, 902 A.2d 888, 896–97 (N.J. 2006). The New Jersey Supreme Court recently expanded the ruling in *Delgado* holding that a hearing on the suggestiveness and reliability of an identification procedure is required in any case in which the police did not record the procedure. *State v. Anthony*, 204 A.3d 229, 241–42 (N.J. 2019).

²⁴⁵ *State v. Romero*, 922 A.2d 693, 703 (N.J. 2007).

²⁴⁶ *State v. Henderson*, 27 A.3d 872, 879–80 (N.J. 2011).

²⁴⁷ *Id.* at 879–83.

²⁴⁸ *Id.* at 884.

²⁴⁹ *Id.* at 884–89.

²⁵⁰ *Id.* at 884, 916.

²⁵¹ *See supra* note 235 (explaining the changes made by the *Henderson* court).

²⁵² Oregon decided *State v. Lawson*, 291 P.3d 673 (Or. 2012), in which it held the same factors articulated in *Henderson* should be applied in Oregon cases, and the trial courts should provide detailed jury instructions where an identification was found admissible. *Id.* at 696–97. In Idaho, the high court decided *State v. Almaraz*, 301 P.3d 242, 252 (Idaho 2013), in which it held that lower courts should apply the *Henderson* factors to determine whether the system and estimator variables resulted in an unreliable identification.

Massachusetts adopted a jury instruction similar to the *Henderson* instruction.²⁵³ Alaska changed its approach to eyewitness identifications in 2016,²⁵⁴ as did Connecticut in 2018²⁵⁵ when it adopted standards mirroring *Henderson*. Additionally, several other jurisdictions, including New York,²⁵⁶ Wisconsin,²⁵⁷ Utah,²⁵⁸ North Carolina,²⁵⁹ District of Columbia,²⁶⁰ Hawaii,²⁶¹ Illinois,²⁶² Maine,²⁶³ Maryland,²⁶⁴ Pennsylvania,²⁶⁵ and Ohio,²⁶⁶ have implemented changes to the way courts evaluate identification procedures, and legislatures have made changes to how police officers conduct identification procedures.²⁶⁷ Other states have acknowledged that issues exist with eyewitness identification procedures, but the courts in those states have not yet adopted significant changes to how these identifications are evaluated.²⁶⁸

²⁵³ Gants & Doughty, *supra* note 242, at 1625–26.

²⁵⁴ Young v. State, 374 P.3d 395, 399 (Alaska 2016).

²⁵⁵ State v. Harris, 191 A.3d 119, 142–43 (Conn. 2018). In Connecticut, similar to New Jersey, its high court reviewed identification issues, including the issues *Henderson* raised, several times before it ultimately adopted the *Henderson* standards. See State v. Johnson, 94 A.3d 1173 (Conn. 2014); State v. Guilbert, 49 A.3d 705 (Conn. 2012). Again, incremental change in the courts led to the more substantial overhaul.

²⁵⁶ See, e.g., People v. Boone, 91 N.E.3d 1194, 1196 (N.Y. 2017).

²⁵⁷ Nicholas A. Kahn-Fogel, *The Promises and Pitfalls of State Eyewitness Identification Reforms*, 104 Ky. L.J. 99, 138–46 (2016).

²⁵⁸ *Id.* at 146–49.

²⁵⁹ *Id.* at 150–57.

²⁶⁰ *In re L.C.*, 92 A.3d 290, 295–96, 301 (D.C. Ct. App. 2014) (holding that trial court should have permitted defendant to call an expert of eyewitness identifications because, as *Henderson* illustrated, reliability of identifications is beyond the ken of the average juror).

²⁶¹ State v. Kaneaiakala, 450 P.3d 761, 777 (Haw. 2019).

²⁶² Though the Illinois courts have not adopted *Henderson*, they have acknowledged that eyewitness identifications are problematic and have permitted expert testimony. See People v. Lerma, 47 N.E.3d 985, 997 (Ill. 2016).

²⁶³ State v. Mahmoud, 147 A.3d 833, 838–39 (Me. 2016).

²⁶⁴ Small v. State, 211 A.3d 236, 246–47 (Md. 2019) (holding that while the court was not adopting *Henderson*, courts can consider the *Henderson* factors as part of the analysis of the admissibility of an identification).

²⁶⁵ Commonwealth v. Walker, 92 A.3d 766, 781, 792–93 (Pa. 2014) (permitting experts in eyewitness identification because of the social science research as articulated in *Henderson*).

²⁶⁶ Kahn-Fogel, *supra* note 257, at 150–57.

²⁶⁷ See generally *id.* at 120.

²⁶⁸ See, e.g., State v. Plain, 898 N.W.2d 801 (Iowa 2017); People v. Sanchez, 941 N.Y.S.2d 599 (N.Y. App. Div. 2016); State v. Davis, 131 A.3d 679 (R.I. 2016).

This is not to say that *Henderson* fixed how all courts review eyewitness identifications. The federal courts and many states still apply the *Manson v. Brathwaite* test,²⁶⁹ and of those states that have changed their legal standards, many courts are still reluctant to find identification procedures improper even if they did not follow the recommended standards.²⁷⁰ Nevertheless, the *Henderson* opinion is less than a decade old and already five states have almost fully adopted the decision, eleven states have substantially changed their approaches to eyewitness identification, and at least three more are aware of the problems and acknowledge that social science raises doubts on the reliability of the evidence. The consensus among the courts is moving toward the belief that eyewitness identification is not always reliable.

The New Jersey Supreme Court is particularly receptive to applying developments in social science to the law, such that this type of decision likely could not have originated from just any jurisdiction. Despite the uniqueness of the New Jersey Supreme Court, jurisdictions throughout the country adopted its approach to eyewitness identifications. The way in which the *Henderson* decision has changed how courts approach eyewitness identification exemplifies the reversibility of the conformity problem: the social science research was there long before *Henderson*, defense attorneys were challenging this evidence long before *Henderson*, but it took *Henderson* for other courts to resist conforming with the norm and change admissibility standards.²⁷¹

In addition, because New Jersey undertook the heavy lift of the multi-day hearings with multiple experts, other courts were able to adopt the research and reasoning without having to conduct the cost-intensive hearings themselves; the record was already developed and one barrier to deviation was removed.

²⁶⁹ See *supra* note 236 (explaining the *Manson v. Brathwaite* test). See, e.g., *State v. Doap Deng Chuol*, 849 N.W.2d 255, 261 (S.D. 2014); *Monroe v. State*, 28 A.3d 418, 431 (Del. 2011) (citing a Delaware state court case that adopts *Manson*); *State v. Gross*, 776 N.E.2d 1061, 1076 (Ohio 2002); *State v. Jones*, 556 N.W.2d 903, 912 (Minn. 1996); *State v. Sanchez*, 288 P.3d 351, 378 (Wash. Ct. App. 2012); *Ibarra v. State*, 11 S.W.3d 189, 195 (Tex. Crim. App. 1999) (referring to the *Biggers* factors). Some courts refer to this test as the *Biggers* test as the factors considered under *Manson* were adopted from *Neil v. Biggers*, 409 U.S. 188, 198–99 (1972).

²⁷⁰ Kahn-Fogel, *supra* note 257, at 102.

²⁷¹ See discussion *supra* note 226. Just as in that study where subjects were willing to resist conformity when just one person agreed with them, courts were able to resist conformity when they saw the New Jersey Supreme Court rule consistent with their own conclusions.

B. District of Columbia Cases

The FTE litigation in the District of Columbia Superior Court and the attendant Court of Appeals decisions demonstrate how one judge's repeated assertions that judges should apply *Daubert* more strictly to forensic science resulted in one of the most restrictive decisions on FTE expert testimony. Though this litigation is still ongoing and developing, there has been substantial movement in the way judges are treating forensic science evidence. In August 2019, Judge Edelman ruled in *United States v. Tibbs* "that the government's expert witness must limit his testimony to a conclusion that, based on his examination of the evidence and the consistency of the class characteristics and microscopic toolmarks, the firearm cannot be excluded as the source of the casing."²⁷² Though the court did not exclude the testimony, the expert will merely be permitted to testify that the recovered firearm cannot be *excluded* from being the source of the casing.

The *Tibbs* decision, however, did not come out of nowhere; it builds upon decisions over the previous years that slowly but steadily created change. In 2011, the D.C. Court of Appeals examined a case in which the defendant requested a *Frye* hearing on the testimony of the FTE examiner, which was denied, and then the defendant failed to object at trial to the examiner testifying that he was "100%" certain of his findings "to the exclusion of all other firearms."²⁷³ On appeal, the Court of Appeals reviewed the testimony regarding certainty and stated, "[W]e will assume, without deciding, that [FTE] experts should not be permitted to testify that they are 100% certain of a match, to the exclusion of all other firearms."²⁷⁴ The Court of Appeals did not decide the issue because the U.S. Attorney's Office had stated that its policy "is to have firearms experts qualify their conclusions to a reasonable degree of scientific certainty," not unqualified certainty.²⁷⁵

This decision, though not explicitly ruling on the issue, opened the door to challenges to FTE testimony, because though the U.S. Attorney's Office insisted its experts would not testify with unqualified certainty,

²⁷² *United States v. Tibbs*, No. 2016 CF1 19431, 2019 D.C. Super. LEXIS 9, at *2 (D.C. Super. Ct. Sept. 5, 2019). This ruling was based on the court's finding that "reliable principles permit a conclusion that a firearm cannot be excluded as the source of a recovered casing or bullet; indeed, this limited conclusion is supported by the reliable principle that firearms leave toolmark impressions on discharged cartridge casings and the reliable method of viewing those impressions under a comparison microscope." *Id.* at *76-77.

²⁷³ *Jones v. United States*, 27 A.3d 1130, 1135 (D.C. 2011) (internal quotation marks omitted).

²⁷⁴ *Id.* at 1139.

²⁷⁵ *Id.* (internal quotation marks omitted).

that is exactly what the expert did in *United States v. Williams*.²⁷⁶ In that case, however, the defense attorney again failed to object to the FTE examiner's testimony that he had no "doubt in [his] mind" that the recovered bullets came from the relevant firearm and provided additional unqualified certainty statements.²⁷⁷ Despite this testimony, which went against the U.S. Attorney's Office's stated policy, the Court of Appeals held that this was not plain error.²⁷⁸

Concurring in this opinion, however, Judge Easterly, citing the NAS Report as well as the Ballistic Imaging Report, among other publications, laid the foundation for challenges to FTE:

[T]here is only one permissible answer to the question left undecided in *Jones* regarding firearms and toolmark examiners' assertions of certainty in their pattern-matching conclusions: the District of Columbia courts should not allow them. It is well established that expert opinion evidence is admissible if 'it will not mislead the jury and will prove useful in understanding the facts in issue.' Certainty statements such as those elicited by the government in this case are misleading and lack any legitimate utility in criminal trials; they express a solid statistical foundation for individualization that does not currently (and may never) exist.²⁷⁹

Judge Easterly goes on to connect these issues with the wrongful convictions that have come from areas of pattern-matching forensic science.²⁸⁰

Shortly after *Williams*, two cases substantially changed the landscape in D.C.: *Gardner v. United States*²⁸¹ and *Motorola, Inc. v. Murray*.²⁸² In *Gardner*, a case where the defense *did* object to unqualified certainty statements from the FTE examiner, the Court of Appeals held that "in this jurisdiction a firearms and toolmark expert may not give an unqualified opinion, or testify with absolute or 100% certainty, that based on ballistics pattern comparison matching a fatal shot was fired from one firearm, to the exclusion of all other firearms."²⁸³ Then, in *Motorola*, the Court of Appeals ruled that the District of Columbia would

²⁷⁶ *Williams v. United States*, 130 A.3d 343, 346 (D.C. 2016).

²⁷⁷ *Id.*

²⁷⁸ *Id.* at 348.

²⁷⁹ *Id.* at 353–54 (Easterly, J., concurring) (citations omitted).

²⁸⁰ *Id.* at 354–55.

²⁸¹ *Gardner v. United States*, 140 A.3d 1172 (D.C. 2016).

²⁸² *Motorola, Inc. v. Murray*, 147 A.3d 751 (D.C. 2016).

²⁸³ *Gardner*, 140 A.3d at 1184. It is worth noting that despite the U.S. Attorney's Office's assertion that it has a policy of not eliciting unqualified statements of certainty during such testimony, the examiner, again, made a statement of unqualified certainty. *Id.* at 1182.

no longer be a *Frye* jurisdiction, but rather a *Daubert* jurisdiction.²⁸⁴ In that matter, Judge Easterly, again in a concurrence, placed the focus on the admission of forensic science evidence and urged trial courts to make use of the NAS Report and PCAST Report in evaluating forensic science in criminal cases.²⁸⁵

Finally, as the motion to exclude or limit the testimony of the FTE examiner in *Tibbs* was pending, the Court of Appeals issued a new decision in *United States v. Williams*.²⁸⁶ Revising its earlier decision based on these subsequent decisions, the Court of Appeals held that the admission of the unqualified testimony regarding the “match” was plain error.²⁸⁷

This series of cases illustrates how a change in a single judge, or a small group of judges, can result in more widespread change. Over the course of several years, Court of Appeals judges began questioning the efficacy of permitting FTE experts to testify with unqualified certainty, ultimately concluding that it was impermissible. That decision, along with the concurrence in *Motorola* urging judges to follow the NAS Report and PCAST Report, created the perfect climate for Judge Edelman to issue his decision in *Tibbs*.

In the District of Columbia, blind acceptance of FTE is no longer the only way in which judges are evaluating this evidence—the high court is pushing trial courts to critically evaluate the proposed experts. As the trajectory of the cases makes clear, though, the changes were made slowly. Not until 2016, seven years after the NAS Report, did the Court of Appeals explicitly prohibit FTE examiners from testifying to unqualified certainty. Regardless, simply by establishing two ways to approach this evidence (the old way and the new way), pressure to conform to the old way has substantially lost influence.

C. District of Massachusetts Cases

In contrast, what looked like promising developments in the District of Massachusetts did not result in any widespread change in that jurisdiction, though they have influenced other jurisdictions and scholars. In 2005 and 2006, Judge Gertner and Judge Saris issued

²⁸⁴ *Murray*, 147 A.3d at 758–59.

²⁸⁵ *Id.* at 759–60 (Easterly, J., concurring).

²⁸⁶ *See Williams v. United States*, 210 A.3d 734, 736 (D.C. 2019) (granting a rehearing because the *Williams* matter was not finalized when the D.C. Court of Appeals rendered the *Motorola* and *Gardner* decisions).

²⁸⁷ *Id.*

decisions within a month of each other criticizing FTE evidence.²⁸⁸ These decisions came before any of the reports detailing the problems with FTE. Both decisions are often cited in law review articles;²⁸⁹ however, since 2005, only seven cases within Massachusetts have cited to *United States v. Green*. Apart from *United States v. Monteiro*, none of the decisions critiqued or limited testimony involving FTE.²⁹⁰ *Monteiro* has only been cited in seven decisions, and none critiqued or limited testimony involving FTE.²⁹¹ *Green* and *Monteiro* were groundbreaking

²⁸⁸ See *United States v. Green*, 405 F. Supp. 2d 104, 104, 106, 109 (D. Mass. 2005) (opinion issued by Judge Gertner); see also *United States v. Monteiro*, 407 F. Supp. 2d 351, 351, 354–55 (D. Mass. 2006) (opinion issued by Judge Saris).

²⁸⁹ See, e.g., Craig M. Cooley, *Forensic Science and Capital Punishment Reform: An "Intellectually Honest" Assessment*, 17 GEO. MASON U. CIV. RTS. L.J. 299 (2007); Brandon L. Garrett & Gregory Mitchell, *The Proficiency of Experts*, 166 U. PA. L. REV. 901, 940 n.165 (2018); Paul C. Giannelli, *Forensic Science: Daubert's Failure*, 68 CASE W. RES. 869, 910–11 (2018); Paul C. Giannelli, *Regulating Crime Laboratories: The Impact of DNA Evidence*, 15 J.L. & POL'Y 59, 92 (2008); Jennifer Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725, 759 (2011). *Green* has been cited in 73 law review articles in total. Citing References—*United States v. Green*, WESTLAW EDGE, <https://1.next.westlaw.com/> (search for 405 F.Supp.2d 104; select Citing References tab; and select secondary sources as the content type) (last visited Oct. 12, 2020). *Monteiro* has been cited in 41 law review articles Citing References—*United States v. Monteiro*, WESTLAW EDGE, <https://1.next.westlaw.com/> (search for 407 F. Supp. 2d 351; select Citing References tab; and select secondary sources as the content type) (last visited Oct. 12, 2020).

²⁹⁰ With the exception of *Monteiro*, no case cited to *Green* in the context of FTE. See *Josiah v. Rodrigues*, No. 16-11986-FDS, 2017 U.S. Dist. LEXIS 44434, at *6 (D. Mass. Mar. 27, 2017) (citing petitioner's use of *Green* in application for leave to obtain further appellate review but rejecting the federal claim where petitioner had failed to exhaust state claims); *Logan v. Gelb*, 52 F. Supp. 3d 122, 135 (D. Mass. 2014) (citing *Green* as an example of a court finding a law enforcement officer was not qualified to testify as an expert); *BASF Corp. v. Sublime Restorations, Inc.*, 880 F. Supp. 2d 205, 212 (D. Mass. 2012) (citing *Green* to quote *Daubert* and *Kumho Tire Co.* in a breach of contract claim); *United States v. Hebshie*, 754 F. Supp. 2d 89, 113–14 (D. Mass. 2010) (finding ineffective assistance of counsel for failure to move for a *Daubert* hearing on any expert witness testimony and citing *Green* for its limitation on the ultimate conclusion of an expert witness); *Tuli v. Brigham & Women's Hosp., Inc.*, 592 F. Supp. 2d 208, 215 (D. Mass. 2009) (permitting an expert to testify without reaching an ultimate conclusion and citing *Green* as an example); *United States v. Frabizio*, 445 F. Supp. 2d 152, 169 (D. Mass. 2006) (citing *Green* in rejecting testimony of an expert whose "conclusions assert a level of certainty unjustified by his methodology and experience"); *United States v. Monteiro*, 407 F. Supp. 2d 351, 364 n.1 (D. Mass. 2006) (see discussion *supra* section IV.C).

²⁹¹ Only one case cited to *Monteiro* in the context of FTE at all. See *United States v. Nascimento*, No. 03-10329-PBS, 2009 U.S. Dist. LEXIS 95355, at *7 (D. Mass. Oct. 9, 2009) (citing *Monteiro* for the principle that FTE expert testimony should only be admitted after a *Daubert* hearing and with proper limitation); *In re Pharm. Indus. Average Wholesale Price Litig.*, 491 F. Supp. 2d 20, 85 (D. Mass. 2007) (citing *Monteiro* in a class action against drug manufacturers to explain the court's gatekeeping function); *Hearts On Fire Co., LLC v. Circa, Inc.*, No. 14-cv-11044-DLC, 2017 U.S. Dist. LEXIS 160872, at *16–17 (D. Mass. Sept. 29, 2017) (citing *Monteiro* to quote *Daubert* and *Kumho Tire Co.* in a breach of contract claim); *Rivera v. City of Worcester*, No. 12-40066-TSH, 2014 U.S.

for their time, but their impact on other judges within the district or state has been minimal—the conformity problem persists in Massachusetts.

The *Green* decision, until the final holding, reads as though the FTE testimony is going to be altogether excluded. Judge Gertner first noted that the expert did not take any notes or photographs of the comparisons he did, and when asked how he decides which marks are considered class versus subclass versus individualized, the FTE examiner admitted that he does not have any studies or databases on which he relies for making such a determination.²⁹² The decision goes on to note a lack of error rate studies, as well as a lack of peer review in the specific instance of this case and the field generally.²⁹³

After pages of a scalding indictment of FTE, Judge Gertner addressed the issue that most plagues these judges: precedent. Even acknowledging that the Court of Appeals for the First Circuit, which is binding on her court, had not ruled on the issue of FTE post-*Daubert*,²⁹⁴ and that trial courts have the discretion to include or exclude this testimony, Judge Gertner found the nonbinding precedent too persuasive. As she noted in her decision, no court, even post-*Daubert*, had excluded FTE evidence.²⁹⁵ In the end, she permitted the FTE witness to testify, but prohibited him from concluding (for which he has no substantiation) that the casings were a match to “the exclusion of all other guns.”²⁹⁶

Shortly thereafter, Judge Saris issued her decision in *State v. Monteiro*.²⁹⁷ *Monteiro* differs from *Green* in that *Monteiro* analyzed FTE and the AFTE method generally, found it a sound methodology, but then found that the expert in that particular case should not be permitted to testify unless and until he complied with the AFTE method. Judge Saris

Dist. LEXIS 82937, at *6–8 (D. Mass. June 17, 2014) (citing *Monteiro* in a hearing to preclude opinion testimony of law enforcement officers to describe the court’s requirement to exercise its gatekeeping function on a case-by-case basis and to quote *Daubert* and *Kumho Tire Co.*); *Sublime Restorations, Inc.*, 880 F. Supp. 2d at 214 (D. Mass. 2012) (quoting *Monteiro* in a breach of contract claim to note that the burden of proof is on the proponent of the evidence); *McGovern v. Brigham & Women’s Hosp.*, 584 F. Supp. 2d 418, 424 (D. Mass. 2008) (citing *Monteiro* in a medical malpractice case to explain the court’s gatekeeping function); *Frabizio*, 445 F. Supp. 2d at 158–59, 162–63 (D. Mass. 2006) (citing *Monteiro* in a *Daubert* hearing about the admissibility of forensic examination of photographic evidence to explain the gatekeeping function and burden of proof).

²⁹² *Green*, 405 F. Supp. 2d at 113, 114 n.19.

²⁹³ *Id.* at 116.

²⁹⁴ *Id.* at 123.

²⁹⁵ *Id.* at 122–23.

²⁹⁶ *Id.* at 124.

²⁹⁷ *United States v. Monteiro*, 407 F. Supp. 2d 351, 351 (D. Mass. 2006).

was less critical in her analysis than Judge Gertner, and though noting that FTE does not have error rates, found that “there is no evidence that the tests are inaccurate or otherwise deficient.”²⁹⁸ Nevertheless, Judge Saris’s decision is notable in that she limited the expert’s testimony (should he come into compliance with the AFTE method): he would not be able to testify to any statistical certainty that the firearm is the source of particular markings on a casing, but he would be able to testify to “a reasonable degree of ballistic certainty.”²⁹⁹

Reading through these decisions, particularly Judge Gertner’s decision, one can see her interest in the area of forensic science and her desire to prevent junk science from getting to the courtroom.³⁰⁰ Then, Judge Saris relied on her colleague’s earlier decision on this very issue before reaching her own conclusion to limit FTE testimony.³⁰¹ However, even with their predispositions to look at FTE testimony with a critical eye, both judges could not overcome the conformity problem: that every prior judicial decision on FTE held this evidence admissible. Judge Gertner stated in her decision:

This reliance on long-standing use of ballistics evidence in the courts is troubling. It runs the risk of ‘grandfathering in irrationality,’ without reexamining it in the light of *Kumho* and *Daubert*. It arguably ignores the mandate of *Daubert*, especially where the courts are relying on pre-*Daubert* acceptance of a given scientific technique.³⁰²

²⁹⁸ *Id.* at 367–68. Note that under *Daubert*, the burden is on the proponent to establish that the method is reliable, not on the opposition to establish that it is not reliable. See *Bourjaily v. United States*, 483 U.S. 171, 176 (1987) (holding that the proponent of evidence has the burden of establishing preliminary facts by a preponderance of the evidence).

²⁹⁹ *Monteiro*, 407 F. Supp. 2d at 372. This distinction appears to have significance to courts as judges perceive this limitation as a true limitation on the witness’s ability to state an opinion. But whether this distinction has an impact on jurors is questionable. See generally, Danielle Weiss & Gerald Laporte, *Uncertainty Ahead: A Shift in How Federal Scientific Experts Can Testify*, NAT’L INST. JUST. J., Apr. 2018, at 71, 73 (discussing a change in DOJ policy that experts should not be permitted to testify to a “degree of scientific certainty” because such a statement can be misleading to jurors and has no basis in science).

³⁰⁰ Judge Gertner had previously issued a similar decision in *United States v. Hines*, 55 F. Supp. 2d 62, 70–71 (D. Mass. 1999), in which she limited the testimony of a handwriting expert and the degree to which that expert could testify about certainty. Thus, she was already interested and engaged in the issues of forensic science and whether such expert testimony should reach a jury post-*Daubert*.

³⁰¹ See *Monteiro*, 407 F. Supp. 2d at 373.

³⁰² *United States v. Green*, 405 F. Supp. 2d 104, 123 (D. Mass. 2005).

This reads as though she is about to strongly limit the expert testimony, but then she only prohibited him from claiming his opinion is “to the exclusion of all other guns.”³⁰³

Judges Gertner and Saris were ahead of the curve, writing years before the comprehensive reports were even released. These two decisions have continued to influence judges, academics, and advocates. Again, demonstrating that though changing one judge’s perception of FTE might not change everything, it is part of a movement to slowly change the norm and confront the conformity problem.

V. CONCLUSION

Since the mid-2000s, multiple reports and exonerations have demonstrated the risks of admitting forensic science without adequately ensuring that the evidence is reliable or sufficiently reliable to go before a jury. Nonetheless, judges have continued to admit this evidence. Even if a judge is cautious enough to craft a limitation on the presentation of evidence, that limitation rarely has a functional meaning to a jury; will a jury really understand the difference between 100% certainty and ballistic certainty? A path toward creating pressure on judges to make real and substantive changes is essential to ensuring that the number of wrongful convictions decreases and that our criminal legal system has integrity.

Many barriers stand between the current norm and substantial change in how judges evaluate forensic science: cognitive biases, decades of bad decisions based on faulty evidence, pressures to conform, strongly held beliefs, and prosecutors providing misleading information. The economic analysis would seem to tilt strongly toward this norm of admitting the evidence never changing. But there are judges in the mold of Judge Skelly Wright, who, for a variety of reasons, will be willing to push beyond these barriers and issue decisions substantially limiting or excluding forensic science evidence that is not foundationally valid. For these judges, the cost of conforming is far more than the cost of deviating.

Not all (or even close to all) judges are willing to deviate from a clearly held norm, but not all judges need to be in order to change the conformity problem. As the *Henderson* case and District of Columbia cases demonstrate, once a single court deviates from the norm, more will follow. The early challengers to conformity serve as anchors, and other judges can then issue their decisions knowing they are not standing alone.

³⁰³ *Id.* at 124.

2021] *FORENSIC SCIENCE AND JUDICIAL CONFORMITY* 643

Why some judges are more likely to become these anchors than others remains elusive. In the context of forensic science, it may be that the judge has a particular curiosity when it comes to evidentiary and forensic science issues, that she is a former defense attorney who has previously made such challenges, or that she is a former prosecutor who wants to ensure that evidence used in criminal trials is without reproach. Whatever the reason, giving judges the tools to critically analyze forensic science is essential; if the judge is unaware of the problem, they will not be part of the change. Thus, judicial trainings, strategic litigation, and advocacy on and publicity of these issues are essential and time sensitive. The continued use of unreliable evidence in criminal cases is against our values. But more importantly, it results in real people spending real time in prisons, away from their families and societies, halting their lives, when they are innocent. This must change.