

The Peter A. Allard School of Law

Allard Research Commons

All Faculty Publications

Allard Faculty Publications

2025

Old Habits Die Hard: Precedent, Psychology, and the Admissibility of Forensic Evidence

Sara Gordon

Allard School of Law at the University of British Columbia, gordon@allard.ubc.ca

Follow this and additional works at: https://commons.allard.ubc.ca/fac_pubs



Part of the [Evidence Commons](#)

Citation Details

S. Gordon, "Old Habits Die Hard: Precedent, Psychology, and the Admissibility of Forensic Evidence" *University of Toronto Law Journal* [forthcoming in 2025].

This Article is brought to you for free and open access by the Allard Faculty Publications at Allard Research Commons. It has been accepted for inclusion in All Faculty Publications by an authorized administrator of Allard Research Commons.

OLD HABITS DIE HARD: PRECEDENT, PSYCHOLOGY, AND THE ADMISSIBILITY OF FORENSIC EVIDENCE

ABSTRACT

Forensic evidence, long considered a cornerstone of criminal justice, has faced increasing scrutiny as recent studies and reports expose significant flaws in its scientific foundation. Techniques such as latent fingerprint analysis, microscopic hair comparison, and ballistics matching, which had been widely accepted for decades, are now being challenged for their lack of empirical validation. Reports by the National Research Council and the President's Council of Advisors on Science and Technology have highlighted the deficiencies in these forensic methods, calling into question their reliability and the weight they are given in courtrooms. Yet despite the growing acknowledgement of widespread issues affecting the reliability and validity of many types of forensic evidence, there are surprisingly few successful challenges to the admissibility of this type of forensic evidence, and when the evidence is challenged, it is often found to be admissible. And while *Daubert* and Rule 702 mandate that expert evidence be based on reliable principles and methods, many courts have failed to rigorously apply these standards, often deferring to precedent rather than conducting a thorough analysis of the scientific validity of forensic techniques.

The article argues that cognitive biases play a significant role in the judicial system's continued acceptance of unreliable forensic evidence. Biases like information cascades, the status quo bias, and the omissions bias can cause judges to favor precedent and established practices, even in the face of new scientific evidence challenging the validity of these forensic methods. The discussion also considers how heuristics, like the bias blind spot, contribute to judges' reluctance to reject long-standing but scientifically flawed forensic techniques. Notwithstanding these challenges, judicial education on scientific standards, greater diversity on the bench, and a heightened awareness of cognitive biases could help mitigate these issues and promote more rigorous evaluation of forensic evidence in the courtroom.

TABLE OF CONTENTS

INTRODUCTION3

I. SCIENCE, *DAUBERT* AND PRECEDENT.....8

 A. The *Daubert* Trilogy9

 B. How Effective has Daubert Been in Keeping Junk Science Out of the Courtroom?11

 i. Junky Science: Judicial Notice & the Role of Precedent15

 ii. *Daubert*: I Do Not Think It Means What You Think It Means.....17

II. WHY IS THIS HAPPENING? MAYBE IT’S BIAS (AND THE BIAS BLINDSPOT).....22

 A. Latent Print Analysis & Information Cascades26

 B. Microscopic Hair Analysis & the Status Quo Bias.....35

 C. Forensic Firearms Analysis & the Omission Bias41

III. HOW CAN WE RESPOND TO THE CONTINUED ADMISSION OF UNRELIABLE FORENSIC EVIDENCE?45

CONCLUSION.....57

INTRODUCTION

*It is revolting to have no better reason for a rule of law than that it was laid down in the time of Henry IV. It is still more revolting if the grounds upon which it was laid down have vanished long since, and the rule simply persists from blind imitation of the past.*¹

*Even the most meticulous, careful, and experienced tarot card reader . . . is still a tarot card reader.*²

On February 3, 1974, a 75-year-old woman named Lovey Benovsky was found dead in her home in Torrance California. During the autopsy, the coroner observed “an elliptical laceration of the nose,” which he identified as teeth marks.³ Police soon arrested Walter Marx, a local man who rented a room from Benovsky, on suspicion of murder. A search warrant was later obtained to secure Marx’s fingerprints and a dentist made impressions of his upper and lower teeth. At Marx’s trial for involuntary manslaughter, three dentists testified on behalf of the state that there was a “match” between Marx’s dentition and the victim’s wounds.⁴

The prevailing standard for admitting expert testimony at the time was the “general acceptance” test, which the DC Court of Appeals articulated in 1923 in *Frye v United States*, and the California Supreme Court adopted in *Huntingdon v Crowley*.⁵ Under this test, the *Marx* court was obligated to review the admissibility of the “forensic odontology” testimony in the case by considering whether the identification technique used by the dentists had received general acceptance by recognized experts in the field.

Decided in 1974, *People v. Marx* was the first reported case to consider the admissibility of bite mark comparison evidence in human flesh. In upholding the lower court’s decision to admit the prosecution’s evidence, the court conceded at the outset that “there is no established science of identifying persons from bite marks,” but found that the dentists in this case did apply established techniques—including X-rays, microscopy, and photography—to the solution of a novel problem.⁶ Moreover, the expert testimony reflected an “enthusiastic response to a rare opportunity to develop or extend forensic dentistry into the area of bite mark identification.”⁷

¹ Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1897).

² Radley Balko, Devil in the Grooves, *The Case Against Forensic Firearms Analysis*, (May 25, 2023), <https://radleybalko.substack.com/p/devil-in-the-grooves-the-case-against>.

³ *People v. Marx*, 54 Cal. App. 3d 101, 104 (1975).

⁴ *People v. Marx*, 54 Cal. App. 3d 101, 106 (1975).

⁵ *Huntingdon v. Crowley*, 414 P.2d 382 (1966).

⁶ *People v. Marx*, 54 Cal. App. 3d 101, 107 (1975). The court seems to have conflated this novel forensic technique with the well-established technique of identifying human remains using x-rays and dental records. As in *Marx*, other courts have treated the two techniques somewhat interchangeably, “further insulating bite mark evidence from judicial scrutiny.” M. Chris Fabricant & Tucker Carrington, *The Shifted Paradigm: Forensic Science’s Overdue Evolution from Magic to Law*, 4 VA. J. CRIM. L. 1, 38 (2016).

⁷ The field of forensic odontology also benefited from serial killer Ted Bundy, who brutally raped and murdered at least 30 women in the 1970s. During his attack, Bundy bit one of the sorority house victims on the buttocks and that bite mark was used as evidence at his murder trial. “Because of the high-profile nature of this case, bite mark analysis went “viral” after this conviction and their use in the court system skyrocketed.” Anthony Cardoza, *Forensic Odontology and Bite Mark Analysis: Understanding the Debate*, 51 J. CALIF. DENTAL ASSOC. 1, 1 (2023).

The *Marx* court explained that while “orderly experimentation” may never be possible in the field of forensic dentistry, the trial court judge was able to view models, photographs, and x-rays of both the victim’s wounds and the defendant’s teeth and could therefore see for himself “the extent to which the purported bite marks appear to conform generally to obvious irregularities in defendant’s teeth.”⁸ Given these readily observable similarities, the court declined to apply the *Frye* test to the admissibility of the bitemark evidence, finding instead that the court “did not have to sacrifice its independence and common sense in evaluating” the conclusions reached by the experts in the case.⁹

Because it is now widely recognized to be junk science, bitemark evidence provides a useful illustration of how faulty forensic evidence can gain a foothold in the courtroom, creating controlling precedent that supports its continued admission despite any scientific basis for the technique. In the years that followed, the *Marx* decision “came to be read as a global warrant to admit bite mark identification evidence whenever a person displaying apparent credentials chose to testify to an identification.”¹⁰ Indeed, jurisdictions that have adopted the *Marx* “eyeball test,” have allowed proponents of bite mark analysis, in one author’s words, “to have their cake and eat it too.”¹¹ Courts around the country permitted experts to testify as to the “superior trustworthiness of the scientific bitemark approach”¹² but the scientific basis of the evidence was not subjected to *Frye* scrutiny, or later, to *Daubert* challenges.¹³

And in this way, a forensic technique with no scientific validity—one that has been said to represent an area of “tremendous concern”¹⁴—became an established scientific approach, responsible for dozens of wrongful convictions, but not yet ruled inadmissible by any court.¹⁵ As recently as 2020, the New Jersey Court of Appeals upheld the admission of bitemark evidence, noting that “courts have criticized the discipline and overturned convictions based on DNA

⁸ *People v. Marx*, 54 Cal. App. 3d 101, 111 (1975).

⁹ *People v. Marx*, 54 Cal. App. 3d 101, 111 (1975).

¹⁰ D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on The Dock?* 64 ALBANY L.R. 99, 138 (2000); see also Radley Balko, *It Literally Started With a Witch Hunt: A History of Bite Mark Evidence*, Wash. Post, Feb. 17, 2015, <https://www.washingtonpost.com/news/the-watch/wp/2015/02/17/it-literally-started-with-a-witch-hunt-a-history-of-bite-mark-evidence/>.

¹¹ M. Chris Fabricant & Tucker Carrington, *The Shifted Paradigm: Forensic Science’s Overdue Evolution from Magic to Law*, 4 VA. J. CRIM. L. 1, 54 (2016).

¹² *People v. Slone*, 143 Cal. Rptr. 61, 69 (Cal. App. 2d Dist. 1978); see also *People v. Watson*, 142 Cal. Rptr. 134, 143 (Cal. App. 1st Dist. 1977) (“The principles governing admission of “bite mark” evidence have been authoritatively stated in *People v. Marx*.”); *U. S. v. Martin*, 9 M.J. 731, 738 (N.C.M.R. 1979) (“The scientific acceptance and reliability of the methods used have been recognized by the appellate courts of various states.”); *State v. Swinton*, 847 A.2d 921, 934 (Conn. 2004) (On appeal, the defendant argues that bite mark evidence, in general, is unreliable, and that its acceptance as evidence by the judicial system is “controversial.” We disagree.”).

¹³ M. Chris Fabricant & Tucker Carrington, *The Shifted Paradigm: Forensic Science’s Overdue Evolution from Magic to Law*, 4 VA. J. CRIM. L. 1, 54 (2016).

¹⁴ Texas Forensic Science Commission, Forensic Bitemark Comparison Complaint 12 (April 12, 2016), <https://www.txcourts.gov/media/1454500/finalbitemarkreport.pdf>.

¹⁵ On July 2, 2024, the US Supreme Court unanimously declined to review the case of Charles McCrory, who was convicted in Alabama in 1985 for the murder of his wife, based almost exclusively on bogus bite-mark testimony. *McCrory v. Alabama*, 144 S. Ct. 2483 (2024); see also Liliana Segura & Jordan Smith, *There’s a Junk Science Crisis in Criminal Convictions. Sonia Sotomayor Calls it Out in Alabama Bite-Mark Case*, The Intercept (July 10, 2024), <https://theintercept.com/2024/07/10/bite-mark-supreme-court-sotomayor/>.

evidence or repudiated testimony, but have not overruled cases allowing admission.”¹⁶ In reaching its decision, the court reviewed multiple governmental reports addressing the lack of scientific validity and unreliability of bitemark evidence, as well as a series of published scientific articles and a brief submitted by the Innocence Project concluding that “every neutral scientific body to have examined bitemarks has rejected it as entirely unreliable.”¹⁷ Notwithstanding this body of evidence, however, the New Jersey court ultimately concluded that “bitemark evidence is currently admissible in New Jersey” and the trial court did not err in failing to conduct an evidentiary hearing on the reliability of the bitemark identification evidence.¹⁸

The New Jersey Court of Appeals decision is not unique. 50 years after the court’s decision in *Marx*, as of July 2024, the National Registry of Exonerations lists 33 cases in which forensic odontology or dental evidence was false or misleading forensic evidence that led to a wrongful conviction.¹⁹ Those 33 cases are just a small fraction of the 3,565 people who since 1989 have been juridically determined to be factually innocent of crimes for which they were wrongfully convicted.²⁰ Of these wrongful convictions, 883, or just over 25% were convicted in cases that involved either false or misleading forensic science.

According to the National Institute of Justice, forensic science is a term that includes a wide variety of activities, some of which lack a “well-developed research base, are not informed by scientific knowledge, or are not developed within the culture of science.”²¹ One common forensic science technique involves “pattern analysis,” where examiners compare two patterns or samples and determine whether the samples “match.” And like the bitemark matching evidence in *Marx*, many of the established “forensic feature comparison methods” techniques used in those cases—methods used to compare and match to criminal defendants things like hair and fiber microscopy, ballistics, shoe prints, tire treads, and blood spatter evidence—are derived from methods based largely on the “training and experience” of the examiner, not on objective science.²²

¹⁶ *State v. Fortin*, 234 A.3d 372, 389 (App. Div. 2020).

¹⁷ *State v. Fortin*, 234 A.3d 372, 389 (App. Div. 2020).

¹⁸ *State v. Fortin*, 234 A.3d 372, 389 (App. Div. 2020). The court did note that Fortin was not convicted solely on the bitemark evidence and “the state’s case against the defendant was “extremely strong in terms of circumstantial evidence,” notably that defendant’s DNA was on the cigarette butt found in the pipe where M.P. was murdered, and defendant was in the area at the time and was scratched.” *Id.* at 385. Notwithstanding this other evidence, however, *Daubert* still requires the trial court to analyze the bitemark evidence for reliability, an obligation the Appellate Court found unnecessary. *See infra*, Part ____.

¹⁹ The Nat’l Registry of Exonerations, Exoneration Detail List, <https://www.law.umich.edu/special/exoneration/Pages/detailist.aspx?View={FAF6EDDB-5A68-4F8F-8A52-2C61F5BF9EA7}&FilterField1=Group&FilterValue1=BM> (last visited Aug. 1, 2024).

²⁰ The Nat’l Registry of Exonerations, <https://www.law.umich.edu/special/exoneration/Pages/about.aspx> (last visited Aug. 1, 2024).

²¹ Nat’l Comm. on Forensic Science, Views Document on Definitions, <https://www.justice.gov/ncfs/file/477836/dl> (last visited Aug. 1, 2024).

²² PCAST REPORT at 60-61. As the PCAST report explains, the term “forensic feature-comparison methods” refers to “the wide variety of methods that aim 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 presence of similar patterns, impressions, features, or characteristics in the sample and the source. Examples include the analyses of DNA, hair, latent fingerprints, firearms and spent ammunition, tool and toolmarks, shoeprints and tire tracks, bitemarks, and handwriting.” *Id.* at 23.

In many wrongful conviction cases, it is DNA—“the one really good form of forensic science”²³—that is used to exonerate the accused. In the early 2000s, following a spate of DNA exonerations and concerned that “faulty forensic science analyses may have contributed to wrongful convictions of innocent people,” Congress directed the National Academy of Science to study the problem and in 2009, the National Research Council (“NRC”) issued its report entitled *Strengthening Forensic Science in the United States: A Path Forward*.²⁴ Ultimately the report concluded that “much forensic evidence—including, for example, bitemarks and firearm and toolmark identifications—is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline.”²⁵ Furthermore, the report expressed concern that most of the techniques the NRC considered were developed in crime laboratories to aid police investigations and “researching their limitations and foundations was never a top priority.”²⁶

Several years later, the President’s Council of Advisors on Science and Technology (“PCAST”) expanded on the NRC report in its own 2016 report that sought to provide “clarity about the scientific standards for the validity and reliability of forensic methods,” and to “evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable.”²⁷ Mindful of its mandate to “ensure the validity of forensic evidence used in the Nation’s legal system,” the PCAST report tied its review of various forensic methods to the requirements of Federal Rule of Evidence 702. In particular, the report looked to foundational validity, which requires that a forensic science method “be shown, based on empirical studies, to be repeatable, reproducible, and accurate,” a requirement that corresponds to Rule 702(c)’s requirement of “reliable principles and methods.”²⁸ With regards to forensic feature-comparison methods, the PCAST report notes that “establishing foundational validity based on empirical evidence is thus a *sine qua non*. Nothing can substitute for it.”²⁹ Like the NRC report, the PCAST report was highly

²³ Hon. Jed Rakoff, *Keynote Address: The Future of Crime Labs and Forensic Science*, 57 HOUS. L. REV. 475, 476 (2020).

²⁴ SEE NAT’L RESEARCH COUNCIL OF THE NAT’L ACADEMIES OF SCIENCES, *STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD* (2009), <http://www.nap.edu/catalog/12589.html> [hereinafter NRC REPORT]; see also PRESIDENT’S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, *FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 2* (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final [hereinafter PCAST REPORT] (noting that “[I]ronically, it was the emergence and maturation of a new forensic science, DNA analysis, in the 1990s that first led to serious questioning of the validity of many of the traditional forensic disciplines. Once DNA analysis became a reliable methodology, the power of the technology—including its ability to analyze small samples and to distinguish between individuals—made it possible not only to identify and convict true perpetrators but also to clear wrongly accused suspects before prosecution and to re-examine a number of past convictions.”).

²⁵ NRC REPORT at 107-08.

²⁶ NRC REPORT at 42. Similarly, the majority of forensic science laboratories in the United States are administered by law enforcement agencies and police departments, NRC REPORT at 183, and many of these labs lack adequate training for examiners. NRC REPORT at 15.

²⁷ PCAST REPORT at 1.

²⁸ PCAST REPORT at 4-5. The report also considered “Validity as applied,” which means that “the method has been reliably applied in practice,” which corresponds to 702(d)’s mandate that an expert “has reliably applied the principles and methods to the facts of the case.”

²⁹ PCAST REPORT at 6 (noting that “neither experience, nor judgment, nor good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) can substitute for actual evidence of foundational validity and reliability.”).

critical of many types of forensic evidence, as well as of the way it continued to be used, noting that “most forensic feature-comparison methods (with the notable exception of DNA analysis) have historically been assumed rather than established to be foundationally valid.”³⁰

Since the NRC and PCAST reports were released, states, scientific organizations, and others have produced similar, often scathing reports.³¹ Yet despite the growing acknowledgement of widespread issues affecting the reliability and validity of forensic feature-comparison methods in the criminal justice system, there are surprisingly few successful challenges to the admissibility of this type of forensic evidence, and when the evidence is challenged, it is often found to be admissible. This is the case when the science underlying the forensic technique has been determined to be unreliable,³² lacking in scientific support,³³ and even when it has “no scientific support.”³⁴

Some have suggested that judges may be reluctant to find this sort of evidence inadmissible because they themselves lack a scientific background and don’t feel comfortable independently assessing the reliability of the proposed evidence.³⁵ In other cases, trial court judges express feeling bound by precedent³⁶ or a fear of being overruled by appellate courts.³⁷ Finally, judges are much more likely to be former prosecutors than former defence attorneys,³⁸ causing some to question whether judges might be more likely to favor the prosecution in admissibility rulings “because they trust and share an inherent predisposition to favouring, and confirming, similar legal approaches on the bench.”³⁹

³⁰ PCAST REPORT at 122

³¹ See e.g., Forensic Science State Commissions and Oversight Bodies—A 2022 Update, <https://forensiccoe.org/private/654825e11c28b>; Texas Forensic Science Commission, Forensic Bitemark Comparison Complaint (April 12, 2016), <https://www.txcourts.gov/media/1454500/finalbitemarkreport.pdf>; National Institute of Science and Technology, Bitemark Evidence, A NIST Scientific Foundation Review (March 2023), <https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8352.pdf>.

³² NRC REPORT at 179 (“The uncertainties associated with bloodstain pattern analysis are enormous.”).

³³ PCAST REPORT at 9 (“Few studies—and no appropriate black-box studies—have been undertaken to study the ability of examiners to accurately identify the source of a bitemark.”); PCAST REPORT at 13 (“PCAST finds that there are no appropriate black-box studies to support the foundational validity of footwear analysis to associate shoeprints with particular shoes based on specific identifying marks. Such associations are unsupported by any meaningful evidence or estimates of their accuracy and thus are not scientifically valid.”).

³⁴ NRC REPORT at 161 (“The committee found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA.”).

³⁵ Jed S. Rakoff & Goodwin Liu, *Forensic Science: A Judicial Perspective*, 120 PROCEEDINGS NAT’L ACADEMY SCI. 1, 3 (2023).

³⁶ *State v. Lopez-Martinez*, 256 P.3d 896 (Kan. App. 2010) (“However we find no indication that our Supreme Court is departing from its holding in *Peoples*. Without such an indication, we are compelled to follow *Peoples*.”).

³⁷ See, e.g., *U.S. v. Green*, 405 F. Supp. 2d 104, 109 (“I reluctantly [admit the evidence] because of my confidence that any other decision will be rejected by appellate courts, in light of precedents across the country, regardless of the findings I have made.”).

³⁸ Alliance for Justice, *Broadening The Bench: Professional Diversity And Judicial Nominations* (2014), <https://afj.org/wp-content/uploads/2021/03/AFJ-2014-Professional-Diversity-Report.pdf>; Clark Neily, *Are a Disproportionate Number of Federal Judges Former Government Advocates?* (2021), <https://www.cato.org/study/are-disproportionate-number-federal-judges-former-government-advocates>.

³⁹ Colleen M. Berryessa et al., *Prosecuting From The Bench? Examining Sources Of Pro-Prosecution Bias In Judges*, 28 L. & CRIMINOLOGICAL PSYCHOL. 1, 7 (2023).

As this Article will argue, however, another reason for the significant lack of successful challenges to the admissibility of unreliable forensic techniques—or junk science⁴⁰—may be the unconscious bias of judges in favor of the admission of this type of evidence, even in the face of overwhelming evidence of its unreliability. In reaching this conclusion, Part I of this Article will discuss the interplay between the legal standards that control the admission of forensic evidence—and in particular forensic feature comparison evidence like the bitemark evidence in *Marx*—and the scientific standards that determine the scientific validity of that evidence. This part will look to the role of Federal Rule of Evidence 702 and the Supreme Court’s decision in *Daubert v. Merrell Dow Pharmaceutical Inc.*, both of which govern the admission of scientific testimony in the courtroom. Although judges make decisions about the admissibility of scientific evidence based on those standards, “the overarching subject of the judges’ inquiry is scientific validity.”⁴¹ This part will therefore consider how judges are applying those standards to determine the admissibility of forensic evidence, and the role that precedent plays in the admission of this evidence, especially evidence that does not pass an objective test of scientific validity.

Next, Part II will discuss how cognitive bias may affect the decision-making of judges who admit dubious forensic evidence—with and without applying the *Daubert* factors to the admission of that evidence—and the role that bias and the bias blind spot may play when courts justify their continued admission of flawed forensic science by citing to precedent. In particular, this Part will argue that what appears to be deference to precedent may instead be the impact of various cognitive biases on the judge’s decision-making. This Part will explore several cognitive biases that are particularly relevant to courts’ reliance on precedent in upholding the admission of dubious forensic science—including information cascades, the status quo bias, and the omission bias—and will ultimately conclude that these unconscious biases may have as much to do with the continued admissibility of junk science as the precedential case law judges cite to when admitting it. Finally, Part III considers several ways the courts might respond to the continued admission of unreliable forensic evidence, including increased judicial science education and increased diversity on the bench. This Part also argues that judges should become more aware of their own vulnerability to bias and the bias blind spot as a way to reduce bias, improve decision-making, and promote the more rigorous use of forensic evidence in the courtroom.

I. SCIENCE, *DAUBERT* AND PRECEDENT

It is no surprise that courts have been slow to reject evidence that has been routinely—and often unquestioningly—admitted for decades. Courts are not institutions designed to change quickly and lower courts tend to “follow the doctrinal mandates of higher courts in virtually all cases.”⁴² The law itself is intended to change slowly and the doctrine of *stare decisis*, a Latin term that means “to stand by things decided”—is a bedrock of the American legal system. *Stare decisis* prompts predictability and stability in the law and prevents judges from making arbitrary decisions.⁴³ Under this view, and as Justice Brandeis famously noted, “*Stare decisis* is usually the wise policy, because

⁴⁰ As M. Chris Fabricant, the Innocence Project’s Director of Strategic Litigation defines the term, “‘Junk Science’ sounds like science but there is no empirical basis for the ‘expert opinion’; it is subjective speculation masquerading as science, typically tilted in the government’s favor against an indigent person of color.” M. CHRIS FABRICANT, JUNK SCIENCE AND THE AMERICAN CRIMINAL JUSTICE SYSTEM 26 (2022).

⁴¹ PCAST REPORT at 21.

⁴² Matthew Tokson, *Judicial Resistance and Legal Change*, 82 U. CHICAGO L. REV. 901, 910 (2015).

⁴³ Earl Maltz, *The Nature of Precedent*, 66 N.C. L. REV. 367, 371 (1988).

in most matters it is more important that the applicable rule of law be settled than that it be settled right.”⁴⁴

Science and scientific knowledge, on the other hand, are continually evolving, and often *unsettled*. When new evidence emerges, prior conclusions are reexamined. As the PCAST report notes, the different approaches of law and science “creates an obvious tension, because many courts admit forensic feature-comparison methods based on longstanding precedents that were set before these fundamental problems were discovered.”⁴⁵ In the legal context, therefore, *stare decisis*—or reasoning from precedent—“requires adhering to a prior decision because it is the prior decision, not necessarily because it is correct.”⁴⁶ And while the report declines to try to answer the question of how to resolve this tension from a legal standpoint, it notes that “from a purely scientific standpoint, the resolution is clear. When new facts falsify old assumptions, courts should not be obliged to defer to past precedents: they should look afresh at the scientific issues.”⁴⁷

The following sections will discuss how the legal system’s approach to determining the admissibility of scientific evidence has evolved in response to advances in science. In particular, trial court judges have been assigned a gatekeeping role to ensure expert testimony about forensic feature-comparison methods meets scientific standards for scientific validity. Therefore, while judges make admissibility decisions based on the legal standards, they are also required to consider the underlying scientific validity of the proposed evidence. As this section will also discuss, this task has proven challenging, and *Daubert* has not always been successful in excluding unreliable “junk science” from the courtroom. As the hundreds of people who have been wrongfully convicted based on false or misleading forensic evidence make clear, the “lack of rigor in the assessment of the scientific validity of forensic evidence is not just a hypothetical problem but a real and significant weakness in the judicial system”⁴⁸

A. *The Daubert Trilogy*

In a series of cases that became known as the *Daubert* trilogy, the Supreme Court clarified the role of Rule 702 and assigned judges a gatekeeping role before any kind of scientific evidence could be admitted at trial. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the Court held that judges must examine not only the expert’s methodology, but also its scientific merit when determining admissibility.⁴⁹ A few years later in *General Electric v Joiner*, the Court held that appellate courts should review trial court *Daubert* decisions deferentially, using an abuse of discretion standard.⁵⁰ And in *Kumho Tire Co. v. Carmichael*, the Court extended *Daubert* scrutiny to non-scientific expert evidence.⁵¹

Under *Daubert*, courts are required to assess the proffered evidence to ensure that the evidence itself is “derived by the scientific method,” and that the “testimony’s underlying reasoning or

⁴⁴ *Burnet v. Coronado Oil*, 285 U.S. 393, 406 (1932) (Brandeis, J., dissenting) (emphasis added).

⁴⁵ PCAST REPORT at 143.

⁴⁶ Goutam U. Jois, *Stare Decisis in Cognitive Error*, 75 BROOK. L. REV. 63, 96s (2009).

⁴⁷ PCAST REPORT at 144.

⁴⁸ PCAST REPORT at 22.

⁴⁹ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

⁵⁰ *General Electric v Joiner*, 522 U.S. 136 (1997).

⁵¹ *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).

methodology is scientifically valid.”⁵² In making this determination, *Daubert* instructs the trial court judge to consider whether expert evidence is the product of a method that was generally accepted—the lone requirement under the old *Frye* standard that existed prior to *Daubert*—but also whether that method has been tested and peer reviewed, whether it has a low error rate, and whether there are standards controlling the techniques application.⁵³

While these factors were ostensibly mere interpretations of the applicable Federal Rules of Evidence, Rule 702 was amended in 2000 to require that expert testimony be based, among other things, on “reliable principles and methods” that have been “reliably applied” to the facts of the case,⁵⁴ essentially codifying the *Daubert* decision. And notwithstanding the additional factors judges may consider under *Daubert*, it is clear that admission of forensic evidence “entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid.”⁵⁵ It is the reliability of the methods that produced the testimony itself that therefore determines reliability.⁵⁶

Daubert was meant to be a flexible test, one that embodies a more liberal standard of admissibility for expert opinions than *Frye*, but this flexibility was not “a license to scrutinize sloppily or not at all.”⁵⁷ Moreover, the *Daubert* opinion makes clear that the requirements of 702 were not meant to “apply specially or exclusively to unconventional evidence”⁵⁸ and even evidence that has long been admitted is still subject to *Daubert* scrutiny.⁵⁹ Finally, while appellate courts should give significant deference to trial courts on *Daubert* decisions, if the lower court does not appropriately apply *Daubert*—or does not apply it at all—that can constitute an abuse of discretion.⁶⁰ So while district courts have “broad discretion” in analyzing reliability, “such discretion does not include the decision ‘to abandon the gatekeeping function.’”⁶¹ Indeed, “the importance of [the] gatekeeping function cannot be overstated.”⁶²

⁵² *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593 (1993).

⁵³ See Fed. R. Evid. 702.

⁵⁴ Fed. R. Evid. 702.

⁵⁵ *Daubert*, 509 U.S. at 592–93.

⁵⁶ See Daniel A. Krauss et al., *The Admissibility of Expert Testimony in the United States, the Commonwealth, and Elsewhere*, in 2 PSYCHOLOGICAL EXPERTISE IN COURT: PSYCHOLOGY IN THE COURTROOM 1, 8 (Daniel A. Krauss & Joel D. Lieberman eds., 2009) (“[W]hile the *Daubert* standard is commonly referred to as a reliability standard by legal commentators, it is actually meant to be an examination of the scientific validity of expert testimony.”).

⁵⁷ Michael J. Saks & David L. Faigman, *Expert Evidence After Daubert*, 1 AN. REV. L. & SOCIAL SCI. 105, 108 (2005).

⁵⁸ See *Daubert*, 509 U.S. at 592 n. 11 (noting that “[a]lthough the decision itself focused exclusively on ‘novel’ scientific techniques, we do not read the requirements of Rule 702 to apply specially or exclusively to unconventional evidence. Of course, well-established propositions are less likely to be challenged than those that are novel, and they are more handily defended.”).

⁵⁹ See Michael J. Saks & David L. Faigman, *Expert Evidence After Daubert*, 1 AN. REV. L. & SOCIAL SCI. 105 118 (2005).

⁶⁰ *Kumho Tire*, 526 U.S. at 159 (J. Scalia concurring) (noting that “the *Daubert* factors are not holy writ, in a particular case the failure to apply one or another of them may be unreasonable and hence an abuse of discretion”).

⁶¹ *Nease v. Ford Motor Co.*, 848 F.3d 219, 230 (4th Cir. 2017) (quoting *Kumho Tire*, 526 U.S. at 158–59, 119 S.Ct. 1167 (Scalia, J., concurring)).

⁶² *United States v. Barton*, 909 F.3d 1323, 1331 (11th Cir. 2018); see also *Sardis v. Overhead Door Corp.*, 10 F.4th 268, 284 (4th Cir. 2021) (“[w]e confirm once again the indispensable nature of district courts’ Rule 702 gatekeeping function in all cases in which expert testimony is challenged on relevance and/or reliability grounds.”).

B. How Effective has Daubert Been in Keeping Junk Science Out of the Courtroom?

Notwithstanding this judicial mandate to scrutinize all purported scientific evidence for reliability, *Daubert* and Rule 702 have been remarkably ineffective in keeping so-called “junk science” out of the courtroom. Under *Daubert*, the gatekeeping role of screening out unreliable evidence, which under *Frye* had been primarily left to the experts themselves, is now assigned to the trial judge.⁶³ Notwithstanding the Supreme Court’s confidence in the ability of trial judges to perform this screening function,⁶⁴ however, some lower courts initially expressed concern that they should fill this role. Writing on remand in *Daubert*, for example, Judge Kozinski of the Ninth Circuit describes this “daunting” task as follows:

Our responsibility, then, unless we badly misread the Supreme Court’s opinion, is to resolve disputes among respected, well-credentialed scientists about matters squarely within their expertise, in areas where there is no scientific consensus as to what is and what is not “good science,” and occasionally to reject such expert testimony because it was not “derived by the scientific method.” Mindful of our position in the hierarchy of the federal judiciary, we take a deep breath and proceed with this heady task.⁶⁵

This fear may not have been completely unfounded. As the 2009 NRC report noted, a review of reported judicial opinions reveals that “at least in criminal cases, forensic science evidence is not routinely scrutinized pursuant to the standard of reliability enunciated in *Daubert*.”⁶⁶ A more recent 2019 study of psychological assessment tools like intelligence tests and risk assessment instruments found that while 90% of the assessment tools offered as expert evidence in legal cases have been subjected to empirical testing, only about 67% were generally accepted in the field, and only 40% have favorable psychometric properties.⁶⁷ Despite the questionable validity of many of these tools, however, the authors found that out of 372 legal cases in which these tools were discussed, only 19—or 5.1%—involved a challenge to a tool’s admissibility.⁶⁸ And when challenges did occur, they failed 68% of the time.⁶⁹

⁶³*Daubert*, 509 U.S. at 589 (“[U]nder the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.”); see also *United States v. Bonner*, 648 F.3d 209, 215 (4th Cir. 2011) (“Federal Rule of Evidence 702 appoints trial judges as “gatekeepers of expert testimony” to protect the judicial process from “the potential pitfalls of junk science.”).

⁶⁴ *Daubert*, 509 U.S. at 593 (“We are confident that federal judges possess the capacity to undertake this review.”).

⁶⁵ *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1316 (9th Cir. 1995); see also David L. Faigman, *Mapping the Labyrinth of Scientific Evidence*, 46 HASTINGS L.J. 555, 558 (1995) (questioning whether judges have the time and energy necessary to master complex scientific theories and observing that scientific testimony covers “topics ranging from voice spectrography to gas chromatography, from premenstrual syndrome to post-traumatic stress syndrome, and from identification through bitemarks to identification through handwriting. A judge attempting to swim from shore to shore of this sea would finish, at best, exhausted and, at worst and more likely, drown.”).

⁶⁶ NRC REPORT at 106. While *Daubert* was a federal case, it was quickly adopted by most state courts. See Nat’l Civil Justice Institute, *State-by-State Compendium Standards of Evidence* (as of July 11, 2023), <https://ncji.org/wp-content/uploads/2024/01/Evidence-Standards-by-State-7.12.23.pdf>.

⁶⁷ Tess M. S. Neal et al., *Psychological Assessments In Legal Contexts: Are Courts Keeping “Junk Science” Out Of The Courtroom?* 20 PSYCH. SCI. PUB. INTEREST 135, 135 (2019).

⁶⁸ Tess M. S. Neal et al., *Psychological Assessments In Legal Contexts: Are Courts Keeping “Junk Science” Out Of The Courtroom?* 20 PSYCH. SCI. PUB. INTEREST 135, 152 (2019).

⁶⁹ Tess M. S. Neal et al., *Psychological Assessments In Legal Contexts: Are Courts Keeping “Junk Science” Out Of The Courtroom?* 20 PSYCH. SCI. PUB. INTEREST 135, 152 (2019) (“On those few occasions when challenges did take

Another recent informal study by Judges Jed Rakoff and Goodwin Liu reviewed 300 federal cases involving *Daubert* challenges to the admissibility of scientific evidence and found similar results, particularly in criminal cases. In the civil cases, nearly a quarter of *Daubert* challenges made by either side were successful, but in criminal cases, less than 10% of challenges made by defendants were successful (though the ones made by prosecutors typically succeeded).⁷⁰ Other studies have found that judges often do not engage with or rule on the reliability of forensic evidence⁷¹ or the proficiency of experts.⁷² Moreover, while some do courts refer to the NRC and PCAST reports,⁷³ they have not had an especially significant impact on the admission of dubious forensic evidence.

Given that many types of forensic evidence that fall short of satisfying *Daubert* are still regularly admitted into evidence, it is important to consider the reasons for this. Studies of judicial attitudes and behaviors suggest that judges take their gatekeeping role seriously and endeavor to apply the *Daubert* criteria.⁷⁴ A 2001 study of 400 state court judges concluded that the vast majority of the judges surveyed believed that the role of gatekeeper was an appropriate one.⁷⁵ Moreover, the vast majority of the surveyed judges felt that the *Daubert* factors were helpful in weeding out unreliable

place, they often failed. Only 6 of the 19 cases challenged (32%) succeeded (that is, the psychological assessment evidence was ruled as inadmissible and excluded from evidence 32% of the time challenges were raised.”); *see also* David DeMatteo et al., *Expert Evidence, The (Unfulfilled) Promise of Daubert*, 20 PSYCHOL. SCI. PUBLIC INTEREST 129, 131 (2019) (“Would society look favorably on a medical test, airplane-safety review, or automobile inspection that relied on similarly questionable procedures? Why then should we accept the admission of such questionable psychological assessment tools in our legal system when someone’s liberty or life can be at stake?”).

⁷⁰ Jed S. Rakoff & Goodwin Liu, *Forensic Science: A Judicial Perspective*, 120 PROCEEDINGS NAT’L ACADEMY SCI. 1, 3 (2023).

⁷¹ Brandon L. Garrett & M. Chris Fabricant, *The Myth of the Reliability Test*, 86 FORDHAM L.R. 1559, 1560 (2018) (noting that “even in rulings that do cite to Rule 702, state courts have neglected the critical language concerning reliability in the Rule and have instead reflexively cited precedent and the putative “flexibility” of the Rule to justify the admission of forensic evidence.”).

⁷² Brandon L. Garrett & Gregory Mitchell *The Proficiency of Experts*, 166 U. PA. L. REV. 901, 903 (2018) (noting that courts “routinely accept a witness’s own self-serving statements of expertise buoyed by educational credentials, professional training, or experience, rarely spending much time on this threshold question before moving on to examine the methods used and conclusions reached by the putative expert.”).

⁷³ *Jones v. U.S.*, 27 A.3d 1130, 1137 (D.C. App. 2011) (Appellant urges us to consider recent reports of the National Research Council (2009), and the National Research Council, *Ballistic Imaging* (2008). After considering it, we are still unpersuaded that pattern matching is no longer generally accepted); *see also U.S. v. Otero*, 849 F. Supp. 2d 425, 431 (D.N.J. 2012), *aff’d*, 557 Fed. Appx. 146 (3d Cir. 2014) (unpublished) (In support of this argument, Defendants point to discussions in the 2009 National Research Council’s report entitled *Strengthening Forensic Science in the United States: A Path Forward* (the “NRC Forensic Science Report”) and its 2008 *Ballistics Imaging Report*, each of which called into question the validity of the assumptions about toolmarks that underlie firearms identification. This Court expresses no opinion on whether the practice of firearms and toolmark identification constitutes a “scientific” discipline because that is not the question before the Court. Rather, the Court must consider whether the Government’s proffered expert opinion is reliable according to the principles of *Kumho Tire*.)

⁷⁴ See Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 443 (2001); Jennifer L. Groscup, *The Effects Of Daubert On The Admissibility Of Expert Testimony In State And Federal Criminal Cases*, 8 PSYCHOL. PUB POLICY & L. 339 (2002); Henry F. Fradella et al., *The Impact of Daubert on the Admissibility of Behavioral Science Testimony*, 30 PEPP. L. REV. 403 (2003).

⁷⁵ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433,443 (2001). 55% felt that Daubert had a “great deal of value” in keeping junk science out of court rooms, while 91% believed that the gatekeeping role was appropriate for judges to assume.

science.⁷⁶ Finally, almost all of the judges “talked about the gatekeeping role as their fundamental “function.”⁷⁷ At the same time, a 2003 study of appellate court opinions found that while courts “are attempting to follow the mandates of the *Daubert* decision...[the courts] devoted little discussion to the four *Daubert* criteria...and the *Daubert* criteria do not predict appellate admissibility.”⁷⁸

One possible explanation for courts’ failure to rigorously apply the *Daubert* factors is that judges lack understanding of the factors and of scientific reliability in general. As the Supreme Court noted in *Daubert*, expert evidence “can be both powerful and quite misleading because of the difficulty in evaluating it.”⁷⁹ The 2001 study, for example, found that “the extent to which judges understand and can properly apply the criteria when assessing the validity and reliability of proffered scientific evidence was questionable at best.” In particular, judges had difficulty defining and applying the *Daubert* criteria: While 82% of judges understood general acceptance,⁸⁰ and 71% understood the meaning of peer review and publication,⁸¹ only 6% understood testing⁸² and 4% understood the concept of error rate.⁸³ As the authors note, “[j]udges’ lack of sophistication regarding the scientific meaning of the *Daubert* criteria should cause concern about whether judges are making accurate and reliable assessments of proffered scientific evidence using the *Daubert* criteria”⁸⁴

⁷⁶ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 444 (2001). 88% felt testing was helpful in determining the merits of proffered scientific evidence; 91% felt a known error rate was helpful; 92% felt peer review and publication were helpful, and 93% felt general acceptance was helpful. *Id.*

⁷⁷ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433,451 (2001) (“It is interesting to note that the few judges who felt the gatekeeping role was inappropriate were not commenting on the gatekeeping role per se, but rather on their perceived lack of ability to perform that role due to a lack of sufficient background and training in scientific methods and principles.”).

⁷⁸ Jennifer L. Groscup, *The Effects Of Daubert On The Admissibility Of Expert Testimony In State And Federal Criminal Cases*, 8 PSYCHOL. PUB POLICY & L. 339, 367 (2002).

⁷⁹ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 595 (1993).

⁸⁰ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 448 (2001) (“In fact, no judge asked for a definition of general acceptance before providing a response.”).

⁸¹ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 447 (2001) (Only 10% of the respondents (n = 39) gave a response that clearly reflected a lack of understanding with respect to the application of peer review and publication in the decision-making process. Five percent of the respondents asked for a definition of “peer review and publication” before providing a response.”).

⁸² Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 444 (2001).

⁸³ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 445 (2001).

⁸⁴ Sophia I. Gatowski et al, *Asking the Gatekeepers: A National Survey Of Judges On Judging Expert Evidence In A Post-Daubert World*, 25 LAW & HUMAN BEHAVIOR 433, 453 (2001); see also Jennifer L. Groscup, *The Effects Of Daubert On The Admissibility Of Expert Testimony In State And Federal Criminal Cases*, 8 PSYCHOL. PUB POLICY & L. 339, 367 (2002) (noting that “[p]oor comprehension of the four *Daubert* criteria and their application may account for the lack of appellate court attention to the four *Daubert* criteria. It is possible that judges would like to apply these criteria in their admission decision making, but they may lack the skill to do so.”); see also Emma Cunliffe & Gary Edmond, *Justice Without Science? Judging The Reliability Of Forensic Science In Canada*, 99 CANADIAN BAR REVIEW 65, 68 (2021) (discussing the case of *R v. Bornyk*, 2017 BCSC 849, where counsel and the trial judge

Courts are not solely responsible for the continued admission of junk science, of course. Attorneys also play a role, as they often fail to challenge the admissibility of expert evidence, as in the study where challenges were only made 5.1% of the time.⁸⁵ Like judges, attorneys may also lack the necessary training to understand expert testimony and some research suggests that lawyers may have similar difficulty in evaluating the quality of expert testimony.⁸⁶ Prosecutors and defense attorneys alike would benefit from additional training on evaluating the scientific reliability of forensic evidence. There are more nefarious reasons, of course, as in the discipline of hair and fiber analysis, which the Justice Department knew as early as 2004 “had weaknesses...and that examiners’ “matches” were often wrong.” Notwithstanding this knowledge, the Justice Department but did not make those findings public, nor did prosecutors notify defendants or their attorneys of the findings in more than half of the 250-plus questioned cases the Justice Department reviewed.⁸⁷

Yet another reason for the continued admissibility of junk science is the state of the science itself. As the PCAST report notes, there has been very limited funding for forensic science research in the United States and “serious peer-reviewed forensic science journals focused on feature-comparison fields remain quite limited.”⁸⁸ There are few opportunities for advanced forensic science education⁸⁹ and people working in the forensic sciences often “lack adequate resources, sound policies, and national support.”⁹⁰ The forensic science community itself also bears some responsibility for promoting the use of good science by courts. While some advances in forensic science disciplines over the last several decades—especially the use of DNA technology—have shown great promise in helping law enforcement identify criminals, in other areas “imprecise or exaggerated expert testimony has sometimes contributed to the admission of erroneous or misleading evidence.”⁹¹

Finally, precedent also plays a role. As Neal et al. note, “if lawyers and experts have always used a particular tool without challenge, then a new challenge is not likely forthcoming.”⁹² At the same time, neither Rule 702 nor *Daubert* require such allegiance to “junk science”⁹³ or to the precedent

seemingly misunderstood key portions of the scientific and statistical fingerprint comparison evidence relied upon by the defence).

⁸⁵ Tess M. S. Neal et al., *Psychological Assessments In Legal Contexts: Are Courts Keeping “Junk Science” Out Of The Courtroom?* 20 PSYCH. SCI. PUB. INTEREST 135, 152 (2019).

⁸⁶ Jacqueline Austin Chorn & Margaret Bull Kovera, *Variations In Reliability And Validity Do Not Influence Judge, Attorney, And Mock Juror Decisions About Psychological Expert Evidence*, 43 L. HUM. BEHAV. 542 (2019); Bradley D. McAuliff, *Can Jurors Recognize Missing Control Groups, Confounds, and Experimenter Bias in Psychological Science?* 33 L. HUM. BEHAV. 247 (2009).

⁸⁷ Spencer S. Hsu, *Convicted Defendants Left Uninformed Of Forensic Flaws Found By Justice Dept.*, WASH. POST (April 16, 2012), https://www.washingtonpost.com/local/crime/convicted-defendants-left-uninformed-of-forensic-flaws-found-by-justice-dept/2012/04/16/gIQAWTcgMT_story.html; *see also infra*, Section ____.

⁸⁸ PCAST REPORT at 32.

⁸⁹ NRC REPORT at 224.

⁹⁰ NRC REPORT at xix.

⁹¹ NRC REPORT at 4.

⁹² Tess M. S. Neal et al., *Psychological Assessments In Legal Contexts: Are Courts Keeping “Junk Science” Out Of The Courtroom?* 20 PSYCH. SCI. PUB. INTEREST 135, 154 (2019).

⁹³ *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 159 (1999) (Scalia, J. concurring) (noting that trial courts are obligated to perform a gatekeeping function and exclude “expertise that is fausse and science that is junky.”).

often cited in favor of its admission. In fact, many courts acknowledge that judges should not simply “grandfather in” evidence that has been generally accepted in the past.⁹⁴

As the next section will demonstrate, however, when *Daubert* challenges to forensic evidence are made, courts often fail to apply *Daubert* and instead take judicial notice of a particular forensic technique, while others ignore *Daubert* and instead rely on pre- and post-*Daubert* precedent in finding a technique admissible. Other courts more vigorously apply Rule 702 to the proffered evidence, but a closer analysis reveals that the application is not as robust as it should be to fulfill the court’s gatekeeping rule. In each of these cases, however, these decisions can create the appearance of post-*Daubert* authority, which is then deferred to by later courts, resulting in a cycle where bad science is institutionalized and later used to support the admission of more bad science.⁹⁵

i. Junky Science: Judicial Notice & the Role of Precedent

Some courts rely on *Daubert* to conclude that a particular theory at issue is “so firmly established as to have attained the status of scientific law,” and therefore properly subject to judicial notice under Rule 201.⁹⁶ Yet while the laws of thermodynamics may not require new proof of their validity or reliability, as the *Daubert* court suggested, the same cannot be said of many types of forensic feature comparison methods. In a 2013 case, for example, the Kentucky Supreme Court declined to hold a *Daubert* hearing and permitted the state’s expert to testify that a hair found on a piece of wood alleged to have been used to kill the victim had the same microscopic characteristics as a hair found on the defendant’s boot, even though the 2009 NRC report found that “linking microscopic hair analysis with particular defendants is highly unreliable.”⁹⁷

Nonetheless, the Kentucky court held that courts may “take judicial notice of those scientific methods or techniques that have achieved the status of scientific reliability, and thus a *Daubert* hearing [was] not required” to admit the microscopic hair evidence, which had been admissible in the state of Kentucky for many years.⁹⁸ Other courts have taken judicial notice of bitemark

⁹⁴ See e.g. *U.S. v. Williams*, 506 F.3d 151, 162 (2d Cir. 2007) (noting that “[e]xpert testimony long assumed reliable before Rule 702 must nonetheless be subject to the careful examination that *Daubert* and *Kumho Tire* require.”); *U.S. v. Green*, 405 F. Supp. 2d 104, 118 (D. Mass. 2005) (noting that district courts are “‘obliged to critically evaluate toolmark and ballistics evidence, even though it has been accepted for years pre-*Kumho*,” because failure to do so “would be equivalent to ‘grandfathering old irrationality.’”); *United States v. Saelee*, 162 F. Supp. 2d 1097, 1105 (D. Alaska 2001) (noting that “[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*.”).

⁹⁵ *U.S. v. Tibbs*, 2019 WL 4359486 (DC Super 2019).

⁹⁶ When a court takes judicial notice of a particular technique, this means its reliability is indisputable. See Fed. R. Evid. 201(b) (limiting judicial notice to a “fact that is not subject to reasonable dispute”).

⁹⁷ NRC REPORT at 161 (noting that “it must be confirmed using mtDNA analysis; microscopic studies alone are of limited probative value. The committee found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA.”); see also PCAST REPORT at 120 (noting that after a review of DOJ documents supporting hair microscopy, “PCAST finds that, based on their methodology and results, the papers described in the DOJ supporting document do not provide a scientific basis for concluding that microscopic hair examination is a valid and reliable process.”).

⁹⁸ *Meskimen v. Com.*, 435 S.W.3d 526, 535 (Ky. 2013); see also *U.S. v. Foster*, 300 F. Supp. 2d 375, 377 (D. Md. 2004) (“Ballistics evidence has been accepted in criminal cases for many years.”); *Jones v. U.S.*, 27 A.3d 1130, 1137 (D.C. App. 2011) (“Pattern matching is not new, and courts in this jurisdiction have long been admitting firearms identifications based on this method.”).

evidence,⁹⁹ blood spatter evidence¹⁰⁰ and toolmark identification evidence,¹⁰¹ despite significant concern about the underlying scientific validity of these techniques.

Other courts seem to simply ignore Rule 702 and instead rely on case law prior to the 2000 amendments, or even prior to *Daubert* itself. In *Schmidt v State*, for example, an Iowa appellate court in 2017 declined to consider evidence from the 2009 NRC report that “[t]he uncertainties associated with bloodstain pattern analysis are enormous,” observing that “the admissibility of such evidence is determined by the Iowa Rules of Evidence, not a research journal.”¹⁰² In support of this decision, the court notes that blood pattern evidence has been admissible in Iowa trial courts since 1980, and supports this observation with citations to ten different state and federal court decisions upholding the admissibility of blood pattern evidence, nearly all of which were decided pre-*Daubert*.¹⁰³ Ultimately, the court held that Schmidt’s trial counsel was not ineffective for failing to object to the evidence and did not breach any duty by declining to pursue “this meritless issue.”¹⁰⁴ Similarly, a Kansas Appellate Court in 2010 acknowledged “a 1999 study which reported a 63% rate of misidentifications in bite-mark analyses,” but was “compelled to follow *Peoples*,” a 1980 case in which the Kansas Supreme Court ruled that “[B]ite-mark identification by an expert witness is sufficiently reliable and can be a valuable aid to a jury in understanding and interpreting evidence in a criminal case.”¹⁰⁵

Other courts cite to post-*Daubert* precedent, even though the precedential case relied on science that has since been found to be methodologically unsound. In 2012, the Seventh Circuit in *United States v Smith* upheld the district court’s admission of footwear impression analysis, noting that it had previously upheld the admission of this type of evidence in 2004 in *United States v Allen*.¹⁰⁶ The *Smith* court also cites to decisions from the Third, Sixth, Seventh, and Eighth Circuits, all of which admitted footwear impressions in the early 2000s. In rejecting Smith’s argument that *Allen* should be revisited because footwear-impression analysis does not meet the demands of Rule 702, the court points out that the techniques used in the *Allen* were generally accepted in the forensic community and subject to peer review.¹⁰⁷

In the time since *Allen* was decided in 2004, however, many had begun to question the reliability of footwear impression analysis. According to the 2009 NRC report, for example, there is no consensus in the footwear impression community regarding the number of characteristics needed

⁹⁹ *People v. Marsh*, 441 N.W.2d 33, 36 (Mich. App. 1989).

¹⁰⁰ *Holmes v. State*, 135 S.W.3d 178, 189 (Tex. App.--Waco 2004).

¹⁰¹ *Abuquah v. State*, 296 A.3d 961 (Md. 2023).

¹⁰² *Schmidt v. State*, 200 F.3d 1286, 1297 (10th Cir. 2000).

¹⁰³ *Schmidt v. State*, 200 F.3d 1286, 1297 (10th Cir. 2000) (citing *United States v. Mustafa*, 22 M.J. 165, 168 (C.M.A. 1986); *Robinson v. State*, 574 So. 2d 910, 918 (Ala. Crim. App. 1990); *State v. Downing*, 791 A.2d 649, 654–55 (Conn. App. Ct. 2002); *State v. Rodgers*, 812 P.2d 1208, 1212 (Idaho 1991); *State v. Moore*, 458 N.W.2d 90, 98 (Minn. 1990); *State v. Mix*, 781 P.2d 751, 758 (Mont. 1989); *People v. Murray*, 537 N.Y.S.2d 399, 399 (N.Y. App. Div. 1989); *State v. Goode*, 461 S.E.2d 631, 644 (N.C. 1995); *State v. Deel*, 1986 WL 11203, at *18 (Ohio Ct. App. Sept. 30, 1986); *State v. Melson*, 638 S.W.2d 342, 364 (Tenn. 1982); *Holmes v. State*, 135 S.W.3d 178, 195 (Tex. Crim. App. 2004)).

¹⁰⁴ *Schmidt v. State*, 902 N.W.2d 590 (Iowa Ct. App. 2017).

¹⁰⁵ *State v. Lopez-Martinez*, 256 P.3d 896 (Kan. App. 2010) (citing *State v. Peoples*, 227 Kan. 127, 132, 605 P.2d 135 (1980)).

¹⁰⁶ *United States v. Smith*, 697 F.3d 625, 634 (7th Cir. 2012).

¹⁰⁷ *United States v. Smith*, 697 F.3d 625, 634 (7th Cir. 2012).

to make a positive identification. Moreover, it is difficult to avoid biases in these comparisons and most of the research in the field is conducted in forensic laboratories with results published in trade journals, not peer reviewed journals.¹⁰⁸ Several years later, the PCAST report similarly found that because there are no appropriate studies of footwear matching techniques, “such associations are unsupported by any meaningful evidence or estimates of their accuracy and thus are not scientifically valid.”¹⁰⁹ Despite these findings and Smith’s argument that the court’s previous decision in *Allen* should be revisited because footwear-impression analysis does not meet the demands of Rule 702, the court ultimately “reaffirmed its holding in *Allen*.”¹¹⁰

As many courts have noted, however, *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.¹¹¹ But while some courts have begun to more closely scrutinize forensic evidence long-assumed admissible and to more carefully apply the standards articulated in *Daubert* and codified in Rule 702, the following section will demonstrate that some of those cases still fall short of ensuring the reliability *Daubert* demands of expert forensic evidence.

ii. *Daubert*: I Do Not Think It Means What You Think It Means¹¹²

Even when courts apply *Daubert* and Rule 702 to proposed forensic feature comparison methods, they do not always engage in a robust analysis of the evidence’s reliability or consider *Daubert* factors beyond general acceptance. For example, in 2012 in *United States v Otero*, the government offered the testimony of a Lieutenant of the New Jersey State Police that the spent ammunition recovered from a robbery crime scene was fired from the handguns seized from the defendants at the time of their arrest.¹¹³ In reaching his conclusion, the government’s expert used the Association of Firearm and Tool Mark Examiners (AFTE) theory of identification of firearms, the method commonly used by firearm and toolmark examiners throughout the country.¹¹⁴

The defendants objected to the admission of the evidence and the use of the AFTE method, noting that the 2009 NRC report as well as its 2008 report on the use of ballistics imaging technology called into question the validity of the assumptions about toolmarks that underlie firearms

¹⁰⁸ NRC REPORT at 150.

¹⁰⁹ PCAST REPORT at 13.

¹¹⁰ *U.S. v. Smith*, 697 F.3d 625, 634 (7th Cir. 2012) (“Accordingly, consistent with our holding in *Allen*, we conclude that the district court did not abuse its discretion in admitting FBI Examiner Smith’s expert testimony regarding the shoe print evidence.”).

¹¹¹ Various 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’”; *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]*.”).

¹¹² Change this.

¹¹³ *United States v. Otero*, 849 F. Supp. 2d 425, 428 (3d Cir. 2014).

¹¹⁴ *United States v. Otero*, 849 F. Supp. 2d 425, 428 (3d Cir. 2014).

identification, but the court quickly dismissed those findings.¹¹⁵ First, the court noted that the NRC report “focused on the challenges and limitations faced by a number of forensic science disciplines,” but the court says nothing else about the NRC report, nor about its findings as to firearms and toolmark evidence.¹¹⁶ In fact, the NRC report found that while some individual characteristics of firearms may be distinctive enough to suggest a particular source, it lacked a “precisely defined process” or specific protocol,” and that the best guidance available for the field of toolmark identification . . . does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.”¹¹⁷

Next, the court similarly dismisses the findings of NRC’s 2008 Ballistics imaging report, noting that the report “did not include any commentary on the discipline of firearms identification.”¹¹⁸ This is not entirely accurate, as the NRC Ballistics report did observe that “[t]he validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.”¹¹⁹ Finally, the court similarly dismissed the conclusions of a defense witness, a professor from the John Jay College of Criminal Justice who testified at the trial court that “the discipline is hampered by systemic scientific problems, such as inadequate statistical empirical foundations and a lack of objective standards for identification.”¹²⁰ Having therefore dispensed with the various sources suggesting the unreliability of toolmark evidence as “unpersuasive,” the court proceeded to apply the *Daubert* factors to the government expert’s testimony.¹²¹

In applying the first *Daubert* factor to the proposed testimony, the *Otero* court first concluded that the AFTE theory was testable and had been tested; for this, the court cited to “the many studies demonstrating the uniqueness and reproducibility of firearms toolmarks.”¹²² A closer look at the evidence reveals that two of the three studies the court considered were authored by the same person, Ron Nichols, who owns and operates the Nichols Firearm and Toolmark Identification

¹¹⁵ NRC REPORT at 155 (“This AFTE document, which is the best guidance available for the field of toolmark identification, does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.”); National Academy of Sciences, Ballistics Imaging Report 60 (2008) (“The AFTE theory of identification is rooted in the recognition that “the interpretation of individualization/identification is subjective in nature.” However, it melds that recognition with more objective, quasi-quantitative benchmarks—“sufficient agreement,” “significance,” “likelihood . . . so remote,” and agreement in both “quantity and quality”—but no specific empirical definition is given for these terms.”).

¹¹⁶ *United States v. Otero*, 849 F. Supp. 2d 425, 430 (3d Cir. 2014).

¹¹⁷ NRC REPORT at 155. Shortly after *Orteo* was decided, the PCAST report would observe that because there was “only a single study that was appropriately designed to test foundational validity and estimate reliability . . . firearms analysis currently falls short of the criteria for foundational validity.” PCAST REPORT at 112. And as David Faigman and colleagues would later observe “few studies of firearms exist and those that do indicate that examiners cannot reliably determine whether bullets or cartridges were fired by a particular gun.” David L. Faigman et al., *The Field of Firearms Forensics Is Flawed*, SCIENTIFIC AMERICAN (May 25, 2022), <https://www.scientificamerican.com/article/the-field-of-firearms-forensics-is-flawed/>. As Faigman et al. note, “Contrary to its popular reputation, firearms identification is a field built largely on smoke and mirrors.” *Id.*

¹¹⁸ *United States v. Otero*, 849 F. Supp. 2d 425, 430 (3d Cir. 2014).

¹¹⁹ National Academy of Sciences, Ballistics Imaging Report 3 (2008).

¹²⁰ *United States v. Otero*, 849 F. Supp. 2d 425, 430 (3d Cir. 2014).

¹²¹ *United States v. Otero*, 849 F. Supp. 2d 425, 430 (3d Cir. 2014).

¹²² *United States v. Otero*, 849 F. Supp. 2d 425, 432 (3d Cir. 2014).

Training Academy, and the third was authored by Richard Grzybowski, a firearms examiner with the Bureau of Alcohol Tobacco and Firearms.¹²³

Both Nichols and Grzybowski are experienced firearm and toolmark examiners with many years of experience. But as an influential paper calling for a more robust research culture in the forensic science notes, “most practicing forensic scientists in pattern and impression evidence, and in most other forensic disciplines as well, are not actually qualified to pursue the necessary research.”¹²⁴ Moreover, a significant body of research has demonstrated the many challenges associated with insider research, where research is undertaken within an organization, group, or community where the researcher is also a member.¹²⁵ Specifically, researcher bias can cause the researchers’ personal values and experience to influence the research questions, study methodology, and results.¹²⁶ Because both Nichols and Grzybowski are industry insiders who have a stake in the continued success of the firearm and toolmark identification field, the court’s failure to engage with or cite to any “outsider research” is troubling. Or as Professor Paul Giannelli more bluntly describes it, the field of firearms and toolmarks identification has for years been “an entrenched forensic discipline [that] vigorously guarded its turf by rejecting the conclusions of the outside scientific community.”¹²⁷

The court next considered the second *Daubert* factor of peer review and publication and concluded that the “AFTE theory is subject to peer review through submission to and publication by the AFTE Journal of validation studies which test the theory.”¹²⁸ The AFTE Journal is an industry journal which “publishes articles, studies and reports concerning firearm and toolmark evidence.”¹²⁹ The ATFE Journal, however, like many journals that publish pattern identification research “simply do[es] not comport with broader norms of access, dissemination, or peer review typically associated with scientific publishing.”¹³⁰ Although the journal does claim to subject

¹²³ *United States v. Otero*, 849 F. Supp. 2d 425, 432 (3d Cir. 2014).

¹²⁴ Jennifer L. Mnookin et al, *The Need For A Research Culture In The Forensic Sciences*, 58 UCLA L.R. 725, 759 (2011) (noting that “[e]xperience may provide the basis for determining what questions to ask, but most pattern identification analysts, even with entirely noble intentions, would not be qualified to design or develop sophisticated research projects to answer those questions.”).

¹²⁵ Jenny Fleming, *Recognizing And Resolving The Challenges Of Being An Insider Researcher In Work-Integrated Learning*, 19 INT’L J. WORK-INTEGRATED LEARNING 311, 311 (2018); see also ANDREW LOXLEY & AIDAN, *SOME PHILOSOPHICAL AND OTHER RELATED ISSUES IN INSIDER RESEARCH IN RESEARCHING EDUCATION FROM THE INSIDE* (Pat Sikes & Anthony Potts eds. 2008).

¹²⁶ Jenny Fleming, *Recognizing And Resolving The Challenges Of Being An Insider Researcher In Work-Integrated Learning*, 19 INT’L J. WORK-INTEGRATED LEARNING 311, 313 (2018); Melanie J. Greene, *On the Inside Looking In: Methodological Insights and Challenges in Conducting Qualitative Insider Research*, 19 QUALITATIVE REPORT 1, 4 (2014).

¹²⁷ Paul C. Giannelli, *Forensic Science: Daubert’s Failure*, 68 CASE W. RESRV. L. REV. 869, 916 (2018).

¹²⁸ *U.S. v. Otero*, 849 F. Supp. 2d 425, 433 (3d Cir. 2014).

¹²⁹ *U.S. v. Otero*, 849 F. Supp. 2d 425, 433 (3d Cir. 2014).

¹³⁰ Jennifer L. Mnookin et al, *The Need For A Research Culture In The Forensic Sciences*, 58 UCLA L.R. 725, 754 (2011). As the authors note, “Worldcat—the largest online catalog of library materials, which includes the holdings of 72,000 libraries worldwide, including virtually every university-based library in the United States—lists only eighteen libraries with a copy of this journal in their holdings. Furthermore . . . the only available index to AFTE was created by an individual firearms examiner on his own initiative and was not continued past 2005.” *Id.* at 755; see also Paul C. Giannelli, *Forensic Science: Daubert’s Failure*, 68 CASE W. RESRV. L. REV. 869, 916 (2018). (noting that the ATFE Journal, which is advertised as the “the Scientific Journal” of AFTE, was not generally available until 2016 and is “peer reviewed” by other members of its discipline.).

submissions to peer review, it is unclear where the peer reviewers come from¹³¹ and the editorial board consists entirely of other firearms examiners.¹³² Articles that “that present an opinion, technique, or method having scientific significance are all subject to post-publication peer review by the members of the Association of Firearm & Toolmark Examiners.”¹³³ The majority of the editorial process, therefore, appears to include members of the toolmark and firearms practitioner community, and the journal itself does not appear to be widely disseminated beyond ATFE members.¹³⁴ As Mnookin and colleagues note, “[n]one of this is compatible with an accessible, rigorous, transparent culture of research.”¹³⁵

In considering *Daubert*'s third factor, whether there is a known error rate for the AFTE technique, the Court relies on another article, also published in the AFTE Journal, finding that “while a definitive error rate has not been calculated, the information derived from the proficiency testing is indicative of a low error rate,” typically around 1%.¹³⁶ As David Faigman and colleagues have observed since the court's decision in *Orteo*, however, while studies find persuasively low error rates of around 1 percent, “how the studies arrived at these rates is dubious” and the methodology employed in the studies cited by firearms examiners in court to support these low error rates are often “flawed and seriously misleading.”¹³⁷

Finally, the Court concludes that the “AFTE theory of firearms and toolmark identification is widely accepted in the forensic community and, specifically, in the community of firearm and toolmark examiners.” However, when courts limit the “relevant scientific community” to the practitioners in that forensic science community as the court did in *Orteo*, they inevitably limit that group to practitioners, and not scientists, and “there is a key distinction between this form of expertise and that of a *researcher*, who is professionally trained in experimental design, statistics and the scientific method; who manipulates inputs and measures outputs to confirm that the techniques are valid.”¹³⁸ Or as Radley Balko quipped in a recent article, “a test to assess the

¹³¹ The linked in page of the ATFE journal recently issued a call for additional peer reviewers, though only members of the AFTE are permitted to serve on the AFTE Editorial Committee. See <https://www.linkedin.com/company/firearmtoolmark/posts/?feedView=all>

¹³² AFTE Reviewer Guidance Manual (2015), <https://afte.org/uploads/documents/reviewer-guidance-manual.pdf>. According to the Manual, the AFTE Editorial Committee is composed of Cole Goater (Editor, Nebraska State Patrol Crime Lab firearms expert), Eric Collins (NIBIN Section Chief at Bureau of Alcohol, Tobacco, Firearms and Explosives), Robert Caunt (Vancouver Police Department, Forensic Firearms Unit), Michael Haag (*Forensic Scientist and Shooting Incident Reconstructionist), and William Matty (criminologist with San Bernardino County Sheriff's crime lab). *Id.* at 1.

¹³³ *Id.*

¹³⁴ Jennifer L. Mnookin et al, *The Need For A Research Culture In The Forensic Sciences*, 58 UCLA L.R. 725, 755 (2011). The website of the ATFE journal is itself restricted to ATFE members. See Association of Firearm and Tool Mark Examiners, Journals & Newsletters, <https://afte.org/members/journals-and-newsletters> (last visited Aug. 1, 2024).

¹³⁵ Jennifer L. Mnookin et al, *The Need For A Research Culture In The Forensic Sciences*, 58 UCLA L.R. 725, 756 (2011).

¹³⁶ *United States v. Otero*, 849 F. Supp. 2d 425, 434 (3d Cir. 2014).

¹³⁷ David L. Faigman et al., *The Field of Firearms Forensics Is Flawed*, SCIENTIFIC AMERICAN (May 25, 2022), <https://www.scientificamerican.com/article/the-field-of-firearms-forensics-is-flawed/>.

¹³⁸ David L. Faigman et al., *The Field of Firearms Forensics Is Flawed*, SCIENTIFIC AMERICAN (May 25, 2022), <https://www.scientificamerican.com/article/the-field-of-firearms-forensics-is-flawed/>.

competence of other tarot card readers that was designed and administered by tarot card readers will look quite a bit different than a test administered by skeptical scientists.”¹³⁹

Since *Orteo* was decided and as the field of forensic firearms identification has continued to face growing scrutiny, some courts have become more skeptical of claims made by firearms examiners. Among courts that have taken steps to limit the admissibility of this type of evidence, however, most have simply put limits on the conclusions experts can make about ballistics matching,¹⁴⁰ prohibiting experts from testifying that “there is a match to an exact statistical certainty,”¹⁴¹ or that that a match could be made “to the exclusion of every other firearm in the world.”¹⁴²

Other courts have gone further. A 2022 Maryland court found held that “firearms identification has not been shown to reach reliable results linking a particular unknown bullet to a particular known firearm.”¹⁴³ And in 2023, a Chicago circuit court became the first to bar the use of ballistics matching evidence in a criminal case.¹⁴⁴ In making its ruling, the Chicago court noted that “[t]here are no objective forensic based reasons that firearms identification evidence belongs in any category of forensic science.”¹⁴⁵ The judge then reviewed over a dozen cases in which wrongful convictions resulted from flawed ballistics matching evidence, noting that these cases “should serve as a wake-up call to courts operating as rubber stamps in blindly finding general acceptance” of bullet matching analysis.¹⁴⁶

¹³⁹ Radley Balko, Devil in the Grooves, The Case Against Forensic Firearms Analysis, (May 25, 2023), <https://radleybalko.substack.com/p/devil-in-the-grooves-the-case-against>.

¹⁴⁰ See, e.g., *Williams v. United States*, 210 A.3d 734, 742-43 (D.C. 2019) (holding that trial court erred, not in admitting toolmark examiner’s testimony generally, but rather in allowing the “examiner to provide *unqualified* opinion testimony that purports to identify a specific bullet as having been fired by a specific gun via toolmark pattern matching”); *United States v. Willock*, 696 F.Supp.2d 536, 546–47, 571 (D.Md.2010) (holding that “in light of two recent [NRC] studies ... toolmark examiners must be restricted in the degree of certainty with which they express their opinions”).

¹⁴¹ *U.S. v. Montiero*, 407 F. Supp 351 (D. Mass 2006).

¹⁴² *U.S. v. Green*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005); see also *In re Barrett*, 840 F.3d 1223, 1238 (10th Cir. 2016) (“Ballistics expert Terrance Higgs tied the bullet fragment that killed Eales to Defendant’s .223 Colt H Bar Sporter rifle, ‘to the exclusion of all guns that are made or that will be made.’”); *United States v. Law*, 252 F.3d 1357, 2001 WL 422948, at *1 (5th Cir. 2001) (“[B]allistics expert testified that the cartridge recovered at the earlier robbery and the cartridge used in the Griffin carjacking were used in the same weapon ‘to the exclusion of all other firearms in the world.’”). But see *United States v. Romero-Lobato*, 379 F. Supp. 3d 1111 (D. Nev. 2019) (noting that “the courts that imposed limitations on firearm and toolmark expert testimony were the exception rather than the rule” and that many courts continue to admit “unfettered testimony from firearms examiners.”).

¹⁴³ *Abruquah v. State*, 296 A.3d 961 (Md. 2023); see also C.J. Ciaramella, Maryland Supreme Court Limits Testimony on Bullet-Matching Evidence, Reason (June 22, 2023), <https://reason.com/2023/06/22/maryland-supreme-court-limits-testimony-on-bullet-matching-evidence/>. The AFTE released a statement in response to this decision, expressing disappointment with the court’s ruling and noting that “the vast majority of decisions related to Firearm & Tool Mark Identification have been and continue to be favorable.” Furthermore, the ATFE lamented the fact that the decision “establishes an expectation that trial judges will need to become “amateur scientists” in a manner that we believe is incongruent with both the Maryland Rochkind and United States *Daubert* standards,” a position that appears to undercut the gatekeeping role of the judge in the admission of scientific evidence. The Association of Firearm & Toolmark Examiners, The Association of Firearm & Toolmark Examiners response to *Abruquah v. State of Maryland* Decision (June 23, 2023), <https://afte.org/resources/afte-position-documents>.

¹⁴⁴ *State of Illinois v. Winfield*, 2023 Ill. Cir. Ct. 14066, Revised Order and Memorandum Ruling, https://drive.google.com/file/d/1LeClgcOzly1ATTcoeIDL_KHjSpIfcuRI/view (last visited Aug. 1, 2024).

¹⁴⁵ *Id.* at 36.

¹⁴⁶ *Id.* at 41.

In contrast, while the *Orteo* court did ostensibly engage in a *Daubert* analysis of the proffered ballistics evidence, it took an unacceptably narrow view of the reliability demanded by *Daubert* and Rule 702 and refused to engage with defense evidence suggesting the unreliability of the evidence. The court did nominally apply the *Daubert* standards when it found that the ATFE technique had been tested, shown to have a low error rate, and subjected to peer review, but those findings were almost entirely based on the conclusions of the community of ballistics examiners themselves, validated by studies conducted by those examiners, most of whom were not scientifically trained.

The court also gave far less attention to studies and findings put forth by the defense suggesting that firearms and toolmark evidence does not meet the requirements of *Daubert*. And while the technique did have standards that were largely accepted within the community of other forensic firearms examiners, the court failed to serve the gatekeeping role Rule 702 envisions when it looked mainly to the industry itself to vouch for the reliability of the evidence. So while the court's analysis may have fulfilled the technical requirements of Rule 702, it failed to fulfill its purpose of curtailing the admission of pseudoscientific or unreliable expert testimony.

II. WHY IS THIS HAPPENING? MAYBE IT'S BIAS (AND THE BIAS BLINDSPOT)

Despite the many actors who play a role in preventing the admission of dubious science in the courtroom, judges are the final check on the admission of unreliable junk science under *Daubert* and Rule 702. And as discussed above, it is probably the case that the continued admission of dubious forensic evidence is at least partly attributable to judicial ignorance about what constitutes good science or to deference to precedent beyond what is required by *Joiner*. At the same time, neither of these explanations provide a completely satisfying explanation for the unwillingness of many judges—like the judge in *Orteo*—to even consider new research suggesting that long-admitted forensic techniques may amount to little more than junk science. Indeed, in the many cases where courts decline to even *apply* the *Daubert* factors and instead rely on precedent or take judicial notice of largely debunked forensic techniques, we may need to look beyond these empirically testable explanations to a more subtle and unconscious one: the cognitive biases of judicial decisionmakers. In some instances, it may be that the nonconscious cognitive shortcuts—or heuristics—that all human decisionmakers use to efficiently evaluate information can lead to systematic unconscious bias in judges as they make admissibility decisions about forensic evidence.

Heuristics are mental shortcuts or rules of thumb that simplify decision-making processes, allowing people to make quick judgments without extensive deliberation.¹⁴⁷ Because the world is complex and contains an overwhelming amount of information, heuristics help filter and prioritize information, making it easier for people to focus on the most relevant aspects and ignore less critical details. Common examples include the availability heuristic, where individuals make decisions based on the ease with which examples come to mind,¹⁴⁸ and the anchoring heuristic, which causes individuals to rely heavily on the first piece of information encountered (the anchor)

¹⁴⁷ Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974).

¹⁴⁸ Norbert Schwarz et al., *Ease Of Retrieval As Information: Another Look At The Availability Heuristic*, 61 J. PERSONALITY & SOC. PSYCHOL. 195 (1991).

when making decisions.¹⁴⁹ But while heuristics can be useful, they can also lead to cognitive biases, which are systematic errors in thinking, and decades of research has shown that people exhibit cognitive bias in decision-making regardless of experience, training, or intelligence.¹⁵⁰

All people exhibit cognitive biases in decision-making,¹⁵¹ but people often lack awareness of their own biases. In psychology, naive realism is a cognitive bias wherein people believe they see the world objectively and that people who disagree with them must be uninformed, irrational, or biased.¹⁵² This concept highlights how individuals tend to assume their perceptions and beliefs accurately reflect reality and fail to recognize that other people may interpret the same basic circumstances differently than they do.¹⁵³ At the same time, people can easily see—and often exaggerate—the impact of bias on other people’s decision-making.¹⁵⁴

This difference between how we view our own biases and the extent to which we see it in others is known as the bias blind spot.¹⁵⁵ The vast majority of people are susceptible to the bias blind spot and this tendency appears to be unrelated to intelligence or general cognitive or decision-making ability, nor is it correlated with gender and age.¹⁵⁶ In a 2015 study, only one person out of 661 participants said they were more biased than the average person, although participants varied in the degree to which they thought they were less biased than others.¹⁵⁷ In fact, participants who characterized themselves as less biased than others were more likely to rely on intuition and less accurate in their decision-making.¹⁵⁸ As one researcher noted, “it is not our biases that are our biggest stumbling block; rather it is our biased assumption that we are immune to bias.”¹⁵⁹

¹⁴⁹ Adrian Furnham & Hua Chu Boo, *A Literature Review of the Anchoring Effect*, 40 J. SOCIO-ECON. 35 (2011).

¹⁵⁰ See, e.g., Keith E. Stanovich & Richard F. West, *On the Relative Independence of Thinking Biases and Cognitive Ability*, 94 J. PERSONALITY & SOCIAL PSYCHOL. 672 (2008).

¹⁵¹ Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974); Daniel Kahneman et al., *Judgment under Uncertainty: Heuristics and Biases* (Cambridge Univ. Press 1982); Richard Nisbett & Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Prentice-Hall 1980).

¹⁵² Jennifer K. Robbennolt & Jean R. Sternlight, *Psychology for Lawyers* 16 (2d. ed. 2021).

¹⁵³ In film and storytelling, the concept is often referred to as the “Rashomon effect,” from the 1950 Japanese film “Rashomon,” directed by Akira Kurosawa. In the film, several characters provide contradictory accounts of the same incident, each version being influenced by the character’s personal perspective and motivations. Jennifer K. Robbennolt & Jean R. Sternlight, *Psychology for Lawyers* 33 (2d. ed. 2021).

¹⁵⁴ Emily Pronin, *How We See Ourselves and How We See Others*, 320 SCIENCE 1177, 1178 (2008).

¹⁵⁵ Emily Pronin, *Perception And Misperception Of Bias In Human Judgment*, 11 TRENDS COGNITIVE SCI. 37 (2007); Emily Pronin et al., *The Bias Blind Spot: Perceptions of Bias in Self Versus Others*, 28 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 369 (2002).

¹⁵⁶ Irene Scopelliti, *Bias Blind Spot: Structure, Measurement, and Consequences*, 61 MANAGEMENT SCI. 2468, 2478 (2015) (“Instead, the belief that one is less biased than others appears to reflect a lack of self-awareness rather than a more general inferior decision-making ability.”). In fact, higher cognitive ability has been found to be associated with a larger bias blind spot. Richard F. West et al., *Cognitive Sophistication Does Not Attenuate the Bias Blind Spot*, 103 J. PERSONALITY & SOCIAL PSYCHOL. 506, 515 (2012).

¹⁵⁷ Irene Scopelliti, *Bias Blind Spot: Structure, Measurement, and Consequences*, 61 MANAGEMENT SCI. 2468, 2475 (2015).

¹⁵⁸ Irene Scopelliti, *Bias Blind Spot: Structure, Measurement, and Consequences*, 61 MANAGEMENT SCI. 2468, 2476 (2015).

¹⁵⁹ Cynthia McPherson Frantz, *I AM Being Fair: The Bias Blind Spot as a Stumbling Block to Seeing Both Sides*, 28 BASIC & APPLIED SOC. PSYCHOL. 157, 166 (2006).

Judges are no exception to this general rule, and the literature suggests that judges also unintentionally rely on cognitive biases to nearly the same extent as the general public.¹⁶⁰ For example, various studies have shown that judges are just as prone as other people to the hindsight bias, which is our tendency to overestimate the predictability of past events.¹⁶¹ In the courtroom, this means that when a judge must determine what a party “should have known,” “ex ante estimates of appropriate care to be taken against causing harm are apt to appear unreasonable when judged ex post.”¹⁶² Other studies have shown that judges are just as vulnerable to the anchoring bias as other people.¹⁶³ Finally, like other people, training and experience as a judge does not seem to reduce cognitive bias in judges, or as Edmond and Martire recently observed, “being a judge does not necessarily protect against bias.”¹⁶⁴

Moreover, judges are similarly blind when it comes to identifying those biases in themselves.¹⁶⁵ For example, a 2005 study found that exposure to a civil plaintiff’s prior criminal conviction influenced judges’ decision-making, even when the judge themselves found the evidence inadmissible.¹⁶⁶ In that study, judges who ruled that the prior criminal convictions were not admissible awarded an average of 12% less in pain and suffering damages than did those judges who were not exposed to the plaintiff’s criminal history. Moreover, in a 2009 survey asking Administrative Law Judges to rate their capacity for avoiding racial bias in judging, 97.2 percent placed themselves in the top half—so 50 percent placed themselves in the top quartile and 47.2 percent placed themselves in the second quartile. Not a single judge placed themselves in the bottom quartile.¹⁶⁷

¹⁶⁰ See e.g., Gary Edmond & Kristy A. Martire, *Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making*, 82 MODERN L.R. 603 (2019); Emma Cunliffe, *Judging Fast and Slow: Using Decision-making Theory to Explore Judicial Fact Determination*, 18 INT’L J. OF EVIDENCE & PROOF 139 (2014).

¹⁶¹ Jeffrey Rachlinski, *Heuristics and Biases in the Courts: Ignorance or Adaptation?* 79 OR. L. REV. 61, 69 (2000).

¹⁶² Jeffrey Rachlinski, *Heuristics and Biases in The Courts: Ignorance or Adaptation?* 79 OR. L. REV. 61, 70 (2000).

¹⁶³ See, e.g. Birte Englich et al., *Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts’ Judicial Decision Making*, 32 PERSONALITY & SOC. PSYCHOL. BULLETIN 188, 192 (2006) (finding that “sentencing decisions of experienced legal professionals may indeed be influenced by clearly irrelevant sentencing anchors.”); Jeffrey J. Rachlinski et al., *Can Judges Make Reliable Numeric Judgments? Distorted Damages and Skewed Sentences*, 90 INDIANA L.J. 695, 736 (2015) (finding that “even hopelessly arbitrary anchors affected the decisions of the judges we studied.”).

¹⁶⁴ Gary Edmond & Kristy A. Martire, *Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making*, 82 MODERN L.R. 603 (2019) (citing various studies); see also Andrew Wistrich et al., *Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings?* 93 TEXAS L.R. 855, 911 (2015) (concluding that “[m]ost judges try to faithfully apply the law, even when it leads them to conclusions they dislike, but when the law is unclear, the facts are disputed, or judges possess wide discretion their decisions can be influenced by their feelings about litigants

¹⁶⁵ See e.g., Jeffrey J. Rachlinski et al., *Inside the Bankruptcy Judge’s Mind*, 86 B.U. L.R. 1227 (2006) (“Research on some experts—including doctors, real estate agents, psychologists, auditors, lawyers, and judges—shows that they often make the same kinds of mistakes the rest of us make.”); Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 780 (2001) (noting that “[a] plethora of empirical studies establish that cognitive biases, sometimes including anchoring, infect the judgments of professionals, including doctors, lawyers, accountants, real estate appraisers, option traders, psychologists, military leaders, and engineers.”); Andrew J. Wistrich et al., *Can Judges Ignore Inadmissible Information? The Difficulty Of Deliberately Disregarding*, 153 U. PENN. L.R. 1251 (2005).

¹⁶⁶ Chris Guthrie & Andrew J. Wistrich, *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, 153 UNIV. PENN. L. REV. 1251 (2005).

¹⁶⁷ Chris Guthrie et al., *The “Hidden Judiciary”: An Empirical Examination Of Executive Branch Justice*, 58 DUKE L.J. 1477 (2009).

In the courtroom, naïve realism and the bias blind spot may cause judges to believe their legal interpretations and judgments are objective. In turn, this can affect their ability to consider alternative viewpoints or recognize their own biases, potentially impacting the fairness of their rulings. Both phenomena may have been at play in the United States Supreme Court's decision in *Scott v. Harris*.¹⁶⁸ In 2001, Georgia police clocked 19-year old Victor Harris driving 73 miles per hour in a 55 mph zone. When police attempted to pull Harris over, Harris did not stop and instead led police on a high-speed chase. After pursuing Harris for nearly ten minutes, Officer Timothy Scott rammed his police cruiser into Harris's car; Harris crashed and was severely injured. Four different police officer dash cameras captured portions of the pursuit. Harris brought a 1983 action claiming that Officer Scott's action constituted deadly force, and that the use of such force in this context violated Harris's right to be free from excessive force during a seizure.

After reviewing the video of the car chase, the Eleventh Circuit Court of Appeals held that the ramming of Harris's car could constitute deadly force and that such force was not reasonable "in a high-speed chase based only on a speeding violation and traffic infractions where there was little, if any, actual threat to pedestrians or other motorists, as the roads were mostly empty and Harris remained in control of his vehicle."¹⁶⁹ In describing the dash cam video, the court observed that:

Harris remained in control of his vehicle, slowed for turns and intersections, and typically used his indicators for turns. He did not run any motorists off the road. Nor was he a threat to pedestrians in the shopping center parking lot, which was free from pedestrian and vehicular traffic as the center was closed. Significantly, by the time the parties were back on the highway and Scott rammed Harris, the motorway had been cleared of motorists and pedestrians allegedly because of police blockades of the nearby intersections.¹⁷⁰

The Supreme Court disagreed and reversed the decision of the Eleventh Circuit, holding that Officer Scott's actions were reasonable when he rammed Harris's vehicle to terminate the chase. Remarkably, the Court relied on the same videotape as the Eleventh Circuit in reaching its decision, noting that the tape itself contradicted the story accepted by the Court of Appeals:

Indeed, reading the lower court's opinion, one gets the impression that respondent, rather than fleeing from police, was attempting to pass his driving test. The videotape tells quite a different story. There we see respondent's vehicle racing down narrow, two-lane roads in the dead of night at speeds that are shockingly fast. We see it swerve around more than a dozen other cars, cross the double-yellow line, and force cars traveling in both directions to their respective shoulders to avoid being hit. We see it run multiple red lights and travel for considerable periods of time in the occasional center left-turn-only lane, chased by numerous police cars forced to engage in the same hazardous maneuvers just to keep up. Far from being the cautious and controlled driver the lower court depicts, what we see on the video more closely resembles a Hollywood-style car chase of the most frightening sort, placing police officers and innocent bystanders alike at great risk of serious injury.¹⁷¹

¹⁶⁸ *Scott v. Harris*, 550 U.S. 372, 395 (2007).

¹⁶⁹ *Harris v. Coweta Cnty., Ga.*, 433 F.3d 807, 815 (11th Cir. 2005).

¹⁷⁰ *Harris v. Coweta Cnty., Ga.*, 433 F.3d 807, 815 (11th Cir. 2005).

¹⁷¹ *Scott v. Harris*, 550 U.S. 372, 379–80 (2007).

While the decision is nearly unanimous, the Court notes in response to the sole dissent by Justice Stevens that “we are happy to allow the videotape to speak for itself,” and provides a link to the recording on the Court’s website.¹⁷²

A few years after the decision in *Scott*, Professor Dan Kahan published an article in the *Harvard Law Review* describing a study he and co-authors conducted to take up the Court’s invitation to “see for yourself” by watching the video.¹⁷³ They showed the video to 1350 people; of those approximately 75% agreed and 26% disagreed that the use of deadly force was warranted.¹⁷⁴ Nearly a quarter of participants, however, agreed or disagreed only slightly, perhaps further disproving the Supreme Court’s thesis that the tape “spoke for itself.” And while members of certain groups—including Black people, people with lower income, and people with more education—were more likely to believe that deadly force was justified, it is clear that “when the tape spoke, it did not say the same thing to everyone.”¹⁷⁵

It should come as no surprise that when judges are asked to perform cognitively challenging tasks, often under substantial time pressure, and in cases whose outcome will profoundly impact the lives of the people involved, that they will often unconsciously use heuristics to efficiently reach decisions and that these heuristics can sometimes lead to cognitive bias.¹⁷⁶ Moreover, naïve realism can cause individual viewers, like the majority in *Scott*, to believe that others, if they are objective, will see the tape the same way they do. And bias blindness means those individuals might not *realize* that others see the tape differently than they do.

Given the impact of the bias blind spot on all human decisionmakers, including judges, when courts justify their continued admission of flawed science by citing to precedent, it may be the case that that the judge, whether they realize it or not, “might be relying on precedent as a heuristic—a cognitive shortcut—and not because it yields the desirable result.”¹⁷⁷ In other words, what appears to be deference to precedent may instead be the impact of various cognitive biases on the judge’s decision-making. The following sections will explore several cognitive biases that are particularly relevant to courts’ reliance on precedent in upholding the admission of dubious forensic science—including information cascades, the status quo bias, and the omission bias—and will ultimately conclude that these unconscious biases may have as much to do with the continued admissibility of junk science as the precedential case law judges cite to when admitting it.

A. *Latent Print Analysis & Information Cascades*

¹⁷² Media Sources, U.S. Supreme Court, <https://www.supremecourt.gov/media/media.aspx>.

¹⁷³ Dan M. Kahan et al., *Whose E Whose Eyes Are You Going to Believe? Scott v. Harris and the Perils of Cognitive Illiberalism*, 122 HARV. L. REV. 837, 841 (2009).

¹⁷⁴ *Id.* Members of certain groups were more likely than others to believe that deadly force was justified. Blacks were less likely than whites to believe that deadly force was justified, as were Lower income people, people with more education, unmarried people, Democrats, and people from the Northeast. Interestingly there was no real difference between people who lived in cities and people who didn’t.

¹⁷⁵ Jennifer K. Robbennolt & Jean R. Sternlight, *Psychology for Lawyers* 17 (2d. ed. 2021).

¹⁷⁶ Jennifer K. Robbennolt & Jean R. Sternlight, *Psychology for Lawyers* 83-84 (2d. ed. 2021).

¹⁷⁷ Goutam U. Jois, *Stare Decisis in Cognitive Error*, 75 BROOK. L. REV. 63, 68 (2009).

Research in the social sciences consistently demonstrates that people are influenced by the behavior of others.¹⁷⁸ Voters are influenced by opinion polls,¹⁷⁹ many people judge restaurant quality by the number of other diners present, and an individual's movie rating can be influenced positively or negatively when they view other people's prior ratings.¹⁸⁰ In behavioral economics, this pattern of conforming behavior is called an information cascade, a phenomenon in which individual decisionmakers rely on the actions or decisions of others, especially when they encounter incomplete or ambiguous information.¹⁸¹ When public information—the actions of others—is perceived as strong, individuals may even ignore their own private information and follow the same course of action, regardless of that private knowledge.

Information cascades often occur in the financial markets, when the endorsement of one economic decisionmaker influences the responses and purchases of others.¹⁸² In 2021, Keith Gill, also known as RoaringKitty on Reddit, created an information cascade when he became convinced that GameStop stock was undervalued and shared his belief with others on Twitter.¹⁸³ A group of investors known as r/WallStreetBets who followed Roaring Kitty began buying up the stock, which ultimately skyrocketed from around \$20 in early January 2021 to an all-time high of over \$480 by the end of the month. RoaringKitty created an information cascade that caused the r/WallStreetBets investors to disregard their own private information—in this case “Wall Street investors” who generally expected that GameStop stock would continue to fall—and instead follow RoaringKitty's advice and his “diamond hands,” and hold the stock.¹⁸⁴

Information cascades occur “when it is optimal for an individual, having observed the actions of those ahead of him, to follow the behavior of the proceeding person without regard to his own information.”¹⁸⁵ Indeed, work in behavioral economics has shown that it is often rational for subsequent decisionmakers to disregard their own information and follow others, and like the Reddit investors who followed RoaringKitty's advice and bought GameStop stock when it was

¹⁷⁸ See e.g., Robert B. Cialdini & Noah J. Goldstein, *Social influence: compliance and conformity*, 55 ANNUAL REV. PSYCH. 591 (2004); JOHN C. TURNER, *SOCIAL INFLUENCE* (Thomson Brooks/Cole Publishing Co 2001).

¹⁷⁹ Rafael Huber, *Feb Correlates Of Informational Cascades: Brain Mechanisms Of Social Influence On Belief Updating*, 10 SOC COGNITION AFFECTIVE NEUROSCIENCE 589 (2015).

¹⁸⁰ Young-Jin Lee et al., *Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings*, 61 MANAGEMENT SCIENCE 2241 (2015).

¹⁸¹ Sushil Bikhchandani et al., *Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades*, 12 J. ECON. PERSPECTIVES 151 (1998). The terms “informational cascade” and “herd behavior” are often used interchangeably in the literature, but there is a difference between the two concepts. In particular, “an informational cascade is said to occur when an infinite sequence of individuals ignores their private information when making a decision, whereas herd behavior occurs when an infinite sequence of individuals make an identical decision, not necessarily ignoring their private information.” Boğaçhan Çelen & Shachar Kariv, *Distinguishing Informational Cascades from Herd Behavior in the Laboratory*, 94 AM. ECON. R. 484, 485 (2004); see also Abhijit V. Banerjee, *A Simple Model of Herd Behavior*, 107 QUARTERLY J. ECON. 797 (1992).

¹⁸² Sushil Bikhchandani et al., *Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades*, 12 J. ECON. PERSPECTIVES 151 (1998).

¹⁸³ Emily Stewart, *Dumb Money and What Actually Happened With GameStop, Explained* (Sept. 29, 2023), <https://www.vox.com/money/2023/9/15/23873474/dumb-money-gamestop-stock-keith-gill-melvin-capital-review>.

¹⁸⁴ Emily Stewart, *Dumb Money and What Actually Happened With GameStop, Explained* (Sept. 29, 2023), <https://www.vox.com/money/2023/9/15/23873474/dumb-money-gamestop-stock-keith-gill-melvin-capital-review>.

¹⁸⁵ DAVID HIRSHLEIFER, *THE BLIND LEADING THE BLIND: SOCIAL INFLUENCE, FADS, AND INFORMATIONAL CASCADES* in *THE NEW ECONOMICS OF HUMAN BEHAVIOR* Ch. 12 (Cambridge Univ. Press, Ierulli, K. and Tommasi, M., eds. 1995).

\$20 and then saw huge returns on their investment, information cascades often lead to a desired outcome. Because the Reddit investors had incomplete information about the “true state of the world,” it was rational to ignore their own private information and decide based on what they believed to be more informative public signals from RoaringKitty.¹⁸⁶

In other cases, if a substantial number of initial decisionmakers receive an incorrect private signal and make an incorrect decision, this can lead to a “reverse cascade,” resulting in widespread public misinformation with potentially significant negative consequences.¹⁸⁷ For example, the now decades-long debate about the potential risks of harm due to childhood vaccinations is the result of research—later debunked—conducted by Dr. Andrew Wakefield, which suggested that regular childhood vaccinations could cause autism.¹⁸⁸ Although other researchers quickly cautioned against premature acceptance of Wakefield’s hypothesis, it nonetheless received extensive media attention, thanks in part to its embrace by celebrity Jenny McCarthy, who insisted that vaccines caused her son’s autism.¹⁸⁹

Because many parents did not understand or were not reading the many journal articles and empirical studies that refuted the connection between vaccines and autism, they learned about the danger of vaccines from other parents on the news and on social media,¹⁹⁰ and were “left with anecdotal stories on which to come to an informed decision on their own about this issue.”¹⁹¹ As a result, many parents “simply follow[ed] the informational cascade,” which became so widespread that parents could be seen as “naïve or uncaring if [they did] not express concern or awareness about the issue and go through some sort of decisional process.”¹⁹² Today, nearly thirty years later, many parents of children with autism continue to report symptoms that began shortly after their children received vaccinations as infants and a growing number of parents still choose not to vaccinate their children.¹⁹³

¹⁸⁶ Jonathan E. Alevy, *Information Cascades: Evidence from a Field Experiment with Financial Market Profession*, 62 J. FINANCE 151, 151 (2007).

¹⁸⁷ Rafael Huber, *Neural Correlates Of Informational Cascades: Brain Mechanisms Of Social Influence On Belief Updating*, 10 SOC COGNITION AFFECTIVE NEUROSCIENCE 589, 589 (2015).

¹⁸⁸ Wakefield’s paper “Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, And Pervasive Developmental Disorder In Children” was published in *The Lancet* on February 28, 1998. The study involved 12 children and drew an association between the measles-mumps-rubella (MMR) vaccine and autism. The *Lancet* later retracted the study and Wakefield lost his medical license. Laura Eggertson, *Lancet Retracts 12-Year-Old Article Linking Autism To MMR Vaccines*, 183 CMAJ 199 (2010). Researchers have since repeatedly shown that autism is not caused by vaccinations. See e.g., Meldgaard Madsen, M.D., Et Al., *A Population-Based Study Of Measles, Mumps, And Rubella Vaccination And Autism*, 347 NEW ENGLAND J. MEDICINE 1477 (2002).

¹⁸⁹ Samantha D. Gottlieb, *Vaccine Resistances Reconsidered: Vaccine Skeptics And The Jenny McCarthy Effect*, 11 BIOSOCIETIES 152 (2016).

¹⁹⁰ Jan Hoffman, *How Anti-Vaccine Sentiment Took Hold in the United States*, NY Times (Sept. 23, 2019), <https://www.nytimes.com/2019/09/23/health/anti-vaccination-movement-us.html>. As a pediatrician quoted in the article notes, “Nowhere is that reinforcement more clamorous than on social media. You may only see your pediatrician a few times a year, but you can spend all day on the internet.”)

¹⁹¹ April M. Barton, *Application of Cascade Theory to Online Systems: A Study of Email and Google Cascades*, 10 MINN. J.L. SCI. & TECH. 473, 485 (2009).

¹⁹² April M. Barton, *Application of Cascade Theory to Online Systems: A Study of Email and Google Cascades*, 10 MINN. J.L. SCI. & TECH. 473, 485 (2009) (Indeed, “parents who permitted vaccination were gullible toadies of status quo medicine.”).

¹⁹³ Samantha D. Gottlieb, *Vaccine Resistances Reconsidered: Vaccine Skeptics And The Jenny McCarthy Effect*, 11 BIOSOCIETIES 152 (2016).

In the courtroom, information cascades can be dangerous because they give outsized authority to the first court to consider a particular matter, even though that first court may have been basing its decision on incorrect information, or in the case of some forensic feature comparison evidence, on junk science. Because judges often rely on precedent to guide their decision-making, they may rely on the decision of a higher court to admit unreliable evidence without questioning the underlying science, thus creating a cascade effect. In this way, “a series of like-minded holdings may transmit information from preceding to succeeding judges—information that eventually obviates the need for further inquiry”¹⁹⁴ and this blind adherence to precedent can institutionalize outdated legal doctrines or unreliable forensic science.

Information cascades have likely had some influence in the area of latent fingerprint analysis, a forensic technique that every federal court of appeals except the Second Circuit has considered and found to be admissible since Rule 702 was amended in 2000, typically without “conduct[ing] any meaningful analysis of reliability.”¹⁹⁵ Instead, courts often note that the error rate is “low,” discuss the experience of the particular examiner, and emphasize that fingerprint evidence has long been “generally accepted.”¹⁹⁶ The Tenth Circuit in *U.S. v. Baines*, for example, agreed with the defendant that the “record [did] not show that the [fingerprinting] technique has been subject to testing that would meet all of the standards of science,” but ruled that the evidence was properly admitted because “fingerprint identification has been used extensively by law enforcement agencies all over the world for almost a century.”¹⁹⁷ Other circuits, including the Fourth Circuit, emphasize how “[f]ingerprint identification has been admissible as reliable evidence in criminal trials in this country since at least 1911.”¹⁹⁸

Latent fingerprint analysis does in fact have a long history in criminal cases; it was first proposed for use in criminal identification in the 1800s and has long been considered entirely trustworthy. In 1985, the FBI’s manual on the “science of fingerprints” stated that “of all the methods of identification, fingerprinting alone has proved to be both infallible and feasible.”¹⁹⁹ When forensic DNA first appeared in the 1980s, it was sometimes called “DNA fingerprinting” to suggest that it was as reliable as fingerprinting, which was then viewed as the premier identification science and one that consistently produced irrefutable results.²⁰⁰

This widespread acceptance of the infallibility of fingerprint evidence was not limited to the justice system. In 2003, the head of the FBI’s fingerprint unit was interviewed on the television news show 60 Minutes and said that “the probability of error in fingerprint analysis is zero percent.”²⁰¹

¹⁹⁴ Eric Talley, *Precedential Cascades: An Appeal*, 73 S. CAL. L.R. 87, 91 (1999).

¹⁹⁵ Brandon L. Garrett & M. Chris Fabricant, *The Myth of the Reliability Test*, 86 FORDHAM L.R. 1559, 1570 (2018).

¹⁹⁶ Brandon L. Garrett & M. Chris Fabricant, *The Myth of the Reliability Test*, 86 FORDHAM L.R. 1559, 1570 (2018).

¹⁹⁷ *U.S. v. Baines*, 573 F.3d 979, 990 (10th Cir. 2009) As Garrett & Fabricant note, while the Baines court does mention error rate and reliability, “it actually conducted what amounted to a Frye “general acceptance” analysis.” Brandon L. Garrett & M. Chris Fabricant, *The Myth of the Reliability Test*, 86 FORDHAM L.R. 1559, 1570 (2018).

¹⁹⁸ *U.S. v. Crisp*, 324 F.3d 261, 266 (4th Cir. 2003) (noting that every Circuit to have addressed the admission of fingerprint identification has done so has found such evidence admissible).

¹⁹⁹ Federal Bureau of Investigation, *The Science of Fingerprints: Classification and Uses* iv (Washington, D.C.: U.S. Government Printing Office, 1985).

²⁰⁰ NRC REPORT at 104.

²⁰¹ 60 Minutes, “Fingerprints,” January 5, 2003.

Several months after that segment aired, the FBI publicly acknowledged that when investigating the 2004 Madrid train bombings, its examiners had incorrectly matched a print found at the scene to Oregon attorney Brandon Mayfield, in a high profile example of the very real *fallibility* of latent fingerprint identification.²⁰² Police arrested Mayfield and detained him for two weeks before the FBI realized its mistake and he was released.²⁰³

In any discussion of the reliability of fingerprint identification it is important to distinguish latent prints from known prints. A known print contains fingerprint images of up to ten fingers captured in a controlled setting, such as an arrest or a background check. Because known prints tend to be of high quality, they can be searched automatically and reliably against large databases.²⁰⁴ By contrast, in the criminal context, fingerprint analysts attempt to match prints found at a crime scene (“latent prints”) to the known print of a suspect. They look at features on the latent print—the way ridges start, stop, and flow, for example—and compare those features to those of the suspect print to determine whether there is sufficient similarity between the two to declare a “match.”²⁰⁵ In a typical case, a fingerprint analyst uses a technique known as ACE-V to compare a latent print with a known print using a high-powered microscope.²⁰⁶ Unlike known prints, however, latent prints in criminal cases are often incomplete and of variable quality (smudged or otherwise distorted), with quality and clarity depending on such factors as the surface touched and the mechanics of touch.²⁰⁷

The 2009 NRC report was particularly critical of latent print analysis, concluding that there was no available scientific evidence of the validity of the ACE-V method, but noting that “[t]he courts sometimes appear to assume that fingerprint evidence is irrefutable.”²⁰⁸ Several academic articles published after the Mayfield case also noted the lack of scientific validation of fingerprint identification.²⁰⁹ Several years later, the PCAST report observed that “in response to the 2009 NRC report, the latent print analysis field has made progress in recognizing the need to perform

²⁰² An FBI examiner concluded with “100 percent certainty” that the fingerprint matched Brandon Mayfield, an American in Portland, Oregon. See FBI National Press Office, “Statement on Brandon Mayfield Case,” May 24, 2004, <https://archives.fbi.gov/archives/news/pressrel/press-releases/statement-on-brandon-mayfield-case>; Office of the Inspector General, Oversight and Review Division, A Review of the FBI’s Handling of the Brandon Mayfield Case (U.S. Dep’t of Justice, 2006), <https://oig.justice.gov/sites/default/files/archive/special/s0601/final.pdf>.

²⁰³ Studies of the Mayfield case suggested that confirmation bias was a contributing factor to the erroneous identification, causing FBI agents to be influenced by knowledge about other forensic examiners’ decisions. See e.g., Itiel E. Dror et al., *Cognitive Issues In Fingerprint Analysis: Inter- And Intra-Expert Consistency And The Effect Of A ‘Target’ Comparison*, 208 FORENSIC SCI. INT’L 10, 16 (2011); Paul C. Giannelli, *Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs*, 86 N.C. L. REV. 163, 203 (2007).

²⁰⁴ This is also what made it possible to open your phone with your fingerprint. See Devin Coldeway, *iPhone’s New Fingerprint Sensor Knows You From Adam*, NBC News (Sept. 10, 2013), <https://www.nbcnews.com/technolog/iphones-new-fingerprint-sensor-knows-you-adam-8c11122983>.

²⁰⁵ PCAST REPORT at 88. Some analysts will deviate from this method and look at the latent print alongside the suspect’s print before deciding which characteristics are important.

²⁰⁶ NRC REPORT at 135. The acronym “ACE-V” stands for Analysis, Comparison, Evaluation, and Verification. 9 It has been described in forensic literature as a means of comparative analysis of evidence since 1959. *Id.* at 157.

²⁰⁷ As the NRC report notes, “not all fingerprint evidence is equally good, because the true value of the evidence is determined by the quality of the latent fingerprint image.” NRC REPORT AT 7-8.

²⁰⁸ NRC REPORT AT 103.

²⁰⁹ See, e.g., Jonathan J. Koehler, *Fingerprint Error Rates And Proficiency Tests: What They Are And Why They Matter*, 59 HASTINGS L.J. 1077 (2008); Lyn Haber & Ralph Norman Haber, *Scientific Validation Of Fingerprint Evidence Under Daubert*, 7 J. L. PROBABILITY & RISK 87 (2008); Jennifer L. Mnookin, *The Validity Of Latent Fingerprint Identification: Confessions Of A Fingerprinting Moderate*, 7 J. L. PROBABILITY & RISK 127 (2008).

empirical studies to assess foundational validity and measure reliability.”²¹⁰ In particular, the FBI “led the way in performing black-box studies to assess validity and estimate reliability, as well as so-called “white-box” studies to understand the factors that affect examiners’ decisions.”²¹¹ After reviewing the two relevant FBI studies that had been performed since the NRC report was released, as well as a 2012 study out of the Miami-Dade Police Department Forensic Services Bureau, the PCAST report concluded that many fingerprint examiners can, under some circumstances, “produce correct answers at some level of accuracy.”²¹²

The PCAST report noted, however, that the assessment of latent prints from crime scenes is based largely on subjective human interpretation and therefore vulnerable to bias,²¹³ and that the positive rates shown by the studies are much higher than the public and the courts believe, especially given longstanding claims about the accuracy of fingerprint matching.²¹⁴ Moreover, the report concluded that it would be appropriate to inform jurors of those false positive rates, as well as the fact that at the time of its report, there were only two properly designed studies of the accuracy of latent fingerprint analysis. This, the report suggested, would “appropriately inform jurors that errors occur at detectable frequencies, allowing them to weigh the probative value of the evidence.”²¹⁵

Since the PCAST report was released in 2016, there have been more studies, but there are still many things we don’t know about latent print analysis and “the available science does not enable examiners to prove that only one person could be the source of an unknown print.”²¹⁶ Specifically, there are still no objective standards for declaring a match between latent and known prints and this conclusion is instead left to individual examiners;²¹⁷ unlike a technique like DNA analysis, which can generate a statistical probability of a match, fingerprint examinations “rely on subjective assessments of a match that do not lend themselves to specific probabilities”,²¹⁸ and there are serious concerns about “observer effects” and confirmation bias, where examiners can be influenced by outside or contextual information, as well as their own expectations, to see what they expect to see.²¹⁹

Given these deficiencies, some defendants have sought to introduce testimony challenging the validity and accuracy of the ACE-V method of fingerprint examinations based on the PCAST report—including PCAST findings about error rates and recommended jury instructions—but

²¹⁰ PCAST REPORT at 87.

²¹¹ PCAST REPORT at 132.

²¹² PCAST REPORT at 95.

²¹³ PCAST REPORT at 10.

²¹⁴ The false-positive rate could be as high as 1 error in 306 10 cases based on the FBI study and 1 error in 18 cases based on a study by another crime laboratory. PCAST REPORT at 9-10.

²¹⁵ PCAST REPORT at 96.

²¹⁶ Joseph B. Kadane & Jonathan J. Koehler, *Certainty & Uncertainty in Reporting Fingerprint Evidence*, 147 DAEDALUS 119, 131 (2018).

²¹⁷ Austin Hicklin et al., *Why Do Latent Fingerprint Examiners Differ In Their Conclusions?* 316 FORENSIC SCI. 1, 1 (2020) (explaining that examiners “compare latent (fingerprints or palmprints from crime scenes) to exemplars (prints collected from known subjects) to determine whether the latents can be attributed to specific subjects, making subjective conclusions based on their expertise.”).

²¹⁸ Brandon Garrett et al, *Comparing Categorical and Probabilistic Fingerprint Evidence*, 63 J. FORENSIC SCI. 1712, 1712 (2018).

²¹⁹ Itiel E. Dror et al., *Cognitive Issues In Fingerprint Analysis: Inter- And Intra-Expert Consistency And The Effect Of A ‘Target’ Comparison*, 208 FORENSIC SCI. INT’L 10, 16 (2011).

courts almost uniformly deny these requests.²²⁰ In some cases, courts note that PCAST's conclusions about the reliability of latent fingerprint evidence and the ACE-V method go to weight, not admissibility,²²¹ and can be adequately explored through cross-examination of the government's expert, which is the "traditional and appropriate means of attacking shaky but admissible evidence."²²² Others exclude the evidence altogether, finding that defense expert opinions about the NRC and PCAST reports are "not helpful to the jury,"²²³ or are "overly reliant on hearsay."²²⁴ Still others reject the defence evidence by citing longstanding caselaw admitting fingerprint evidence, as the District Court did in *United States v. Reyes-Ballista* when it concluded that "[d]efendant's generic claims [about the unreliability of the government's fingerprint evidence] dissipate in the face of the overwhelming caselaw standing for the proposition that fingerprint evidence is reliable enough for jury trials as a helpful form of identification testimony."²²⁵

It may be, however, that information cascades are also playing a role in judges' decisions to exclude defense evidence about the PCAST report and the known error rates associated with the

²²⁰ See e.g. *U.S. v. Hendrix*, 2020 WL 30342 (W.D. Wash. Jan. 2, 2020); *U.S. v. Reyes-Ballista*, 2020 WL 6822372 (D.P.R. Nov. 20, 2020); *U.S. v. Bonds*, 922 F.3d 343 (7th Cir. 2019); *U.S. v. Cantoni*, 2021 WL 5829754, at *2 (2d Cir. Dec. 9, 2021); *U.S. v. Casanova*, 886 F.3d 55 (1st Cir. 2018). But see *State v. Johnson*, 195 N.E.3d 164 (Ohio 2022) (holding that the trial court erred by not holding a hearing on defendant's petition for postconviction relief). Johnson claimed that his conviction was void because it was based on false and misleading evidence presented by the state. Specifically, a police officer testified as a fingerprint expert that the science of latent-print comparison was an "unbiased and objective process" free from error, and that she was "100% certain" that the latent print found on a bag belonged to Johnson. The appellate court cited both the NRC and PCAST Reports, including PCAST's suggested jury instruction in latent print cases and remanded for a hearing, suggesting that the expert likely erred by testifying that latent fingerprint identification was not subjective and by testifying to 100 percent certainty. *Id.*

²²¹ See e.g., *U.S. v. Kimble*, 2018 WL 4265349 (S.D. Ga. Aug. 16, 2018) (holding that any criticism of ACE-V methodology goes to weight and not admissibility); *U.S. v. Pitts*, 2018 WL 1116550 (E.D.N.Y. Feb. 26, 2018) (holding that defendant could cross-examine based on the Report and that the Report's findings went to the weight and not admissibility of expert's testimony). But see Advisory Comm. on Evidence Rules, *Agenda for Committee Meeting* 105, 107 (Apr. 30, 2021) (noting that "[U]nfortunately many courts have held that the critical questions of the sufficiency of an expert's basis [for his testimony], and the application of the expert's methodology, are generally questions of weight and not admissibility. These rulings are an incorrect application of Rules 702 and 104(a) and are rejected by this amendment.").

²²² See e.g. *Rodriguez v. State*, 30 A.3d 764 (Del. 2011) (noting that "[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence."); *U.S. v. Bonds*, 922 F.3d 343 (7th Cir. 2019) (holding that defendant's concerns about the Government expert's application of ACE-V method can be explored on cross-examination). But see Gary Edmond et al, *Forensic Science Evidence and the Limits of Cross-Examination*, 42 MELBOURNE UNIV. L.R. 858, 869 (2019) (observing that "rather than operating as an engine for exposing weakness and uncovering truth, the impact of cross-examination is inconsistent and often banal.").

²²³ *U.S. v. Hendrix*, , 2020 WL 30342 (W.D. Wash. Jan. 2, 2020); *U.S. v. Pitts*, 2018 WL 1116550 (E.D.N.Y. Feb. 26, 2018) (Finding that proposed defense expert's testimony based on PCAST Report regarding fingerprint analysis would not be helpful to the trier of fact because (1) it would be offered only to rebut opinions that would not be elicited from the Government's expert; and (2) to the extent that proposed defense expert's opinions largely relied on reports authored by others, defendant could cross the Government's experts about those reports and proposed defense expert's testimony in this regard is therefore unnecessary).

²²⁴ *U.S. v. Hendrix*, 2020 WL 30342 (W.D. Wash. Jan. 2, 2020); *U.S. v. Cantoni* (holding that proposed defense expert cannot testify about studies cited in the PCAST report as that would constitute inadmissible hearsay).

²²⁵ *U.S. v. Reyes-Ballista*, 2020 WL 6822372 (D.P.R. Nov. 20, 2020). Similarly, in *U.S. v. Casaus*, the court denied defendant's *Daubert* challenge to fingerprint comparisons, nothing that it is bound by 10th Circuit precedent that "fingerprint comparison is a reliable method to identify persons" undercuts PCAST Report's criticism of latent fingerprint analysis. 2017 WL 6729619 (D. Colo. Dec. 29, 2017).

ACE-V method of fingerprinting. In *U.S. v Llera-Plaza*, for example, Judge Louis Pollak, a well-respected judge and former professor and dean of Yale Law School and the University of Pennsylvania became the first judge in nearly 100 years to question the reliability of fingerprint evidence when he concluded that “fingerprint identification techniques have not been tested in a manner that could be properly characterized as scientific” and were “hard to square” with *Daubert*.²²⁶

In a lengthy opinion, Judge Pollak found that the ACE-V method of latent fingerprint analysis met only one of the *Daubert* criteria, that of general acceptance.²²⁷ Given this finding, he permitted the parties to present expert fingerprint testimony pointing out similarities and differences between the defendants and the latent prints found at the crime scene, but did not permit the experts to express an opinion that a latent print “matched” that of a particular person.²²⁸ The opinion was a shock to prosecutors and defence attorneys alike and received extensive media attention and commentary from around the country.²²⁹

Six weeks later, however, Judge Pollak changed his mind. In a second opinion, the judge bluntly stated, “I disagree with myself.” Following a second evidentiary hearing, he concluded that the fingerprint identification testimony in the case was admissible, and the fingerprint examiners should be permitted to testify about their opinions on whether the fingerprints matched those of the defendant, despite the technique’s latent defects. Judge Pollak granted the government’s motion for reconsideration of his earlier ruling and his original opinion was vacated and superseded.

Judge Pollak explains that his decision to reconsider his earlier ruling was based not on new facts or controlling law, but because he did not base his first decision on the testimony of live witnesses and instead made his decision based on “a transcript of testimony presented in another courtroom more than two years ago.”²³⁰ Notwithstanding this explanation, however, the judge’s first opinion was quite detailed, devoting significant attention to each of the *Daubert* factors, before ultimately concluding that it failed to meet any of the factors except general acceptance. As Jennifer Mnookin points out, “Judge Pollak’s first opinion was the better one.”²³¹

The likelihood that a cascade will start depends on how individuals weight their own private information as compared with publicly available social information.²³² In *Llera-Plaza*, Judge

²²⁶ *U.S. v. Llera Plaza*, 179 F. Supp. 2d 492, 515 (E.D. Pa. 2002).

²²⁷ *U.S. v. Llera Plaza*, 179 F. Supp. 2d 492, 504 (E.D. Pa. 2002).

²²⁸ *U.S. v. Llera Plaza*, 179 F. Supp. 2d 492, 517–18 (E.D. Pa. 2002).

²²⁹ See e.g., Andy Newman, *Judge Rules Fingerprints Cannot Be Called a Match*, *N.Y. Times* (Jan. 11, 2002), <https://www.nytimes.com/2002/01/11/us/judge-rules-fingerprints-cannot-be-called-a-match.html>; Joann Loviglio, *Fingerprinting Under Fire*, CBS News (Feb. 25, 2002), <https://www.cbsnews.com/news/fingerprinting-under-fire/>.

²³⁰ *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 553 (E.D. Pa. 2002).

²³¹ Jennifer L. Mnookin, *Fingerprints: Not a Gold Standard*, 20 ISSUES SCI. TECH. (Fall 2003), <https://issues.org/mnookin-fingerprints-evidence/>.

²³² Rafael Huber, *Neural Correlates Of Informational Cascades: Brain Mechanisms Of Social Influence On Belief Updating*, 10 SOC COGNITION AFFECTIVE NEUROSCIENCE 589, 589 (2015); Jonathan Alevy et al., *Information Cascades: Evidence From An Experiment With Financial Market Professionals*, 62 J. FINANCE 151, 151 (2007) (noting that when “decision makers have imperfect information about the true state of the world, it can be rational to ignore one’s own private information and make decisions based upon what are believed to be more informative public signals.”).

Pollak seems to have given quite a bit of weight to public information—in this case the actions of England, the House of Lords, and Scotland Yard—in reaching his decision. In fact, while he devoted only minimal time in the second opinion to issues of testing, peer review, and general acceptance, he did spend some time in a “historical note”—one that was not drawn from live testimony—discussing “pioneering fingerprint identification efforts,” dating back to the late 1880s in England.²³³ Ironically, he also described what he learned from a defense expert—a senior English fingerprint specialist—who served at New Scotland Yard for twenty-five years. Specifically, he learned that the ACE-V technique used at New Scotland Yard is “is essentially indistinguishable from the FBI’s ACE-V process, and that this formidably knowledgeable and experienced veteran of the Yard—the legendary and actual source of the systematic and comprehensive utilization of fingerprint identification as an instrument of law enforcement—believes in ACE-V without reservation.”²³⁴

He also seems to rely quite heavily on a new fingerprint identification regime that went into force the previous year in England.²³⁵ He notes that Lord Lester of Herne Hill, who described this new regime “in some detail in the House of Lords on February 25, 2002...has been a good friend of the undersigned for some thirty years.” In a footnote, Judge Pollak describes how he asked Lord Lester about the current state of English fingerprint jurisprudence, and this was when he learned about the English regime that had taken effect the previous year. Lord Lester then “formally address[ed] Judge Pollak’s questions to Her Majesty’s Government in the House of Lords on February 25, 2002,”²³⁶ approximately 3 weeks before the second opinion in *Llera-Plaza* came down. Judge Pollak was then gratified to place on the record in his second opinion, “the research results provided ... to Lord Lester,” in response to his questions in the House of Lords, all of which established “that there is no longer any significant lack of harmony between the FBI’s fingerprint identification standards and those that prevail in English courtrooms.”²³⁷

In the case of latent fingerprint evidence, earlier judicial opinions admitting the evidence both before and after *Daubert* have created a “reverse cascade,” one based on incorrect—or in this case scientifically unsupported—information. Like the parents who “followed the informational cascade” in deciding to refrain from vaccinating their children, Judge Pollak ultimately concluded that “arrangements which are felt to be sufficiently reliable in England, ought likewise to be found sufficiently reliable in the federal courts of the United States.”²³⁸ Judge Pollak perceives that the public information available to him—here, the actions of England, Scotland Yard, and Lord Lester—are superior to his own private information and he therefore follows the same course of action in finding the evidence admissible, notwithstanding the private knowledge he gained from

²³³ *U.S. v. Llera Plaza*, 188 F. Supp. 2d 549, 554 (E.D. Pa. 2002).

²³⁴ *U.S. v. Llera Plaza*, 188 F. Supp. 2d 549, 554 (E.D. Pa. 2002).

²³⁵ Judge Pollak notes that this new legal framework “departs very significantly from the regime I had read about in the *Mitchell* record,” though the framework analyzed in both opinions is the ACE-V method of fingerprinting. *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 576 (E.D. Pa. 2002).

²³⁶ *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 568 (E.D. Pa. 2002) (noting that “[t]his is a method of legal research to which I could cheerfully become accustomed.”).

²³⁷ *U.S. v. Llera Plaza*, 188 F. Supp. 2d 549, 570 (E.D. Pa. 2002).

²³⁸ *U.S. v. Llera Plaza*, 188 F. Supp. 2d 549, 576 (E.D. Pa. 2002) (noting that “[t]he techniques of North American fingerprint identification specialists appear to have reached a level of sophistication paralleling that of their English counterparts.”).

his extensive research into and analysis of the ACE-V method that is reflected in his first opinion in *Llera-Plaza*.

Finally, Judge Pollak’s opinion reflects the overall vulnerability of precedent to informational cascades. Like many cases discussing the admissibility of this forensic technique, the second opinion in *Llera-Plaza* makes several references to the long history of fingerprint identification testimony, noting that “English and American trial courts have accepted fingerprint identification testimony for almost a century.” He describes the first English court to endorse fingerprint identification testimony in 1906,²³⁹ as well as the first American court of last resort to admit such evidence in 1911.²⁴⁰ He refers to fingerprint identification evidence as “[t]he bedrock forensic identifier of the 20th century.” In admitting the government’s fingerprint matching testimony in *Llera-Plaza*, Judge Pollak relied on a century of precedent, none of which satisfies requirements of *Daubert* and 702. So like many before him, he followed an information cascade, one that has effectively “obviate[d] the need for further inquiry” into the validity and reliability of latent print analysis.

Ultimately Judge Pollak concluded that while “further research and intellectual scrutiny into the reliability of fingerprint evidence would be “all to the good,” he would not delay the admission of latent print evidence until researchers made headway on verifying and validating the technique. As the judge noted, “to postpone present in-court utilization of this ‘bedrock forensic identifier’ pending such research would be to make the best the enemy of the good.”²⁴¹

B. *Microscopic Hair Analysis & the Status Quo Bias*

The status quo bias describes our tendency to favor the existing state of affairs over change, even when potential changes could lead to better outcomes.²⁴² As Samuelson and Zeckhauser demonstrated in their landmark study, participants consistently showed a preference for maintaining the current state over switching to new options, even when the new options were objectively better.²⁴³ Research on the status quo bias often uses the term interchangeably with “the endowment effect,” in which “people tend to value goods more when they own them than when they do not.”²⁴⁴ In a famous example of the endowment effect, students given coffee mugs were asked to indicate how much money they would want to sell the mug, while students without mugs were asked how much they would pay for a mug. Researchers found that the mug owners

²³⁹ *Rex v. Castleton*, 3 Cr.App. R. 74 (1906).

²⁴⁰ *People v. Jennings*, 252 Ill. 534, 96 N.E. 1077 (1911).

²⁴¹ *U.S. v. Llera Plaza*, 188 F. Supp. 2d 549, 572 (E.D. Pa. 2002). Interestingly, Judge Pollak attributed this quote to Barry *Scheck* and Peter *Neufeld*, co-founders and special counsel at the Innocence Project at the Benjamin N. Cardozo School of Law.

²⁴² William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7 (1988); Daniel Kahneman & Amos Tversky, *Choices, Values, and Frames*, 39 AM. PSYCH. 341, 348 (1984) (“[L]oss aversion favors stability over change.”).

²⁴³ Daniel Kahneman & Amos Tversky, *Choices, Values, and Frames*, 39 AM. PSYCH. 341, 348 (1984).

²⁴⁴ Russell Korobkin, *The Endowment Effect and Legal Analysis*, 97 NW. U. L. REV. 1227, 1228 (2003); see also Maya Bar-Hillel & Efrat Neter, *Why Are People Reluctant to Exchange Lottery Tickets?* 70 J. PERSONALITY & SOC. PSYCHOL. 17 (1996).

consistently set prices that were much higher than the mug buyers were willing to pay.²⁴⁵ In this way, “[f]oregone gains [were] less painful than perceived losses.”²⁴⁶

Like many other cognitive biases, a preference for the status quo can save time and make decision-making more efficient. Samuelson and Zeckhauser use an example of a colleague who has ordered the same ham and cheese sandwich on rye at a local diner for 26 years.²⁴⁷ Or as President Barack Obama explained his decision to wear only blue or gray suits, “I’m trying to pare down decisions. I don’t want to make decisions about what I’m eating or wearing. Because I have too many other decisions to make.”²⁴⁸ In these examples, the sandwich eater and President Obama appear to have made a rational choice to save time and decision-making capacity. In other cases, however, the status quo bias can be irrational or harmful when a change might improve the current way of doing things. For example, there may be a healthier or less expensive sandwich the colleague could eat, or a light-colored suit might be more comfortable in the summer.²⁴⁹

The status quo bias influences decision-making across various domains, and has been cited to explain everything from patient inertia, which causes patients to have trouble changing their behavior to improve their health,²⁵⁰ to brand loyalty, which causes consumers to continue using the same products from the same companies, even when better alternatives may be available.²⁵¹ In another example, Samuelson & Zeckhauser examined investment strategies adopted by university faculty and found that faculty members were more likely to stick with the status quo in health insurance and investment plans, even when better options were available, and even when faculty members knew their original choice was uninformed.²⁵² As one author notes, [p]eople are psychologically uncomfortable with change and will stick with the current state of affairs, even when it directly conflicts with their preferences.²⁵³

Research has identified two related phenomena that are responsible for our bias for the status quo. The first is the influence of anticipated regret, whereby the status quo bias “functions as a regret-minimization strategy.”²⁵⁴ The literature on decision-making has repeatedly demonstrated that

²⁴⁵ Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5, J. ECON. PERSPECTIVES 193, 195-96 (1991) (“For mugs, the median owner was unwilling to sell for less than \$5.25, while the median buyer was unwilling to pay more than \$2.25-\$2.75.”).

²⁴⁶ Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5, J. ECON. PERSPECTIVES 193, 204 (1991).

²⁴⁷ William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 10 (1988).

²⁴⁸ Michael Lewis, *Obama’s Way*, Vanity Fair (Sept 11, 2012), <https://www.vanityfair.com/news/2012/10/michael-lewis-profile-barack-obama>.

²⁴⁹ *But see* Eric Dodds, *In Defense of Obama’s Tan Suit*, Time (Aug. 29, 2014), <https://time.com/3214633/barack-obama-tan-suit/>.

²⁵⁰ Gaurav Suri et al., *Patient Inertia And The Status Quo Bias: When An Inferior Option Is Preferred*, 24 PSYCHOL. SCI. 1763 (2013).

²⁵¹ Xinping Shi, *Consumer Loyalty Toward Smartphone Brands: The Determining Roles Of Deliberate Inertia And Cognitive Lock-In*, 55 INFORMATION & MANAGEMENT 866 (2018).

²⁵² William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 26 (1988).

²⁵³ Jonathan Breslin, *The Status Quo Bias And Decisions To Withdraw Life-Sustaining Treatment*, 190 CMAJ E265 (2018).

²⁵⁴ Jonathan Breslin, *The Status Quo Bias And Decisions To Withdraw Life-Sustaining Treatment*, 190 CMAJ E265 (2018).

individuals regret decisions to maintain the status quo less than decisions to change.²⁵⁵ For example, a family member may choose to maintain life-sustaining treatment because the decision to withdraw treatment may lead the family member to regret the decision later on and wonder if more could have been done to save their loved one.²⁵⁶ Manetti and colleagues call this status quo effect “one of the most robust phenomena in the regret literature.”²⁵⁷

The second phenomena is the closely-related omissions bias, discussed in more detail below, which is driven by the fear of responsibility for one’s own actions.²⁵⁸ Several authors suggest that the omissions bias can lead people to embrace the status quo in order to reduce their responsibility for the outcomes of their choices.²⁵⁹ People opt for the status quo “to avoid being the direct cause of harm and the perceived greater moral responsibility for being the cause of that harm.”²⁶⁰ For example, parents may choose to put their children at risk of harm from an illness rather than feeling responsible for the harm that might come to the child as a result of a vaccine reaction.²⁶¹

Judges have been shown to be vulnerable to the status quo bias, including a study of bankruptcy judges in which judges’ willingness to approve a reorganization plan differed depending on whether the plan was framed as a gain or a loss.²⁶² In fact, the status quo bias has been used to explain stare decisis itself, which American judges adhere to “because it is an inheritance from English tradition and therefore represents the status quo.”²⁶³ The status quo bias can also cause individual judges to adhere to precedent admitting unreliable forensic evidence, even if new information suggests that the existing precedent may be incorrect or misguided. If judges view precedent—or what we can describe as the legal default—as an entitlement like a coffee mug, the endowment effect suggests that judges will place additional value on that precedent because they already possess or understand the legal default. In turn, the judge “will be less likely to opt out of the default.”²⁶⁴ This preference for the status quo in the courtroom means that forensic evidence

²⁵⁵ Jonathan Baron & Ilana Ritov, *Omission Bias, Individual Differences, And Normality*, 94 *ORGAN. BEHAV. HUM. DECISION PROCESS* 74 (2004); Lucia Mannetti et al., *Who Regrets More After Choosing A Non-Status-Quo Option? Post Decisional Regret Under Need For Cognitive Closure*, 28 *J. ECON. PSYCHOL.* 186 (2007).

²⁵⁶ Jonathan Breslin, *The Status Quo Bias And Decisions To Withdraw Life-Sustaining Treatment*, 190 *CMAJ E265* (2018).

²⁵⁷ Lucia Mannetti et al., *Who Regrets More After Choosing A Non-Status-Quo Option? Post Decisional Regret Under Need For Cognitive Closure*, 28 *J. ECON. PSYCHOL.* 186 (2007).

²⁵⁸ Jonathan Baron & Ilana Ritov, *Omission Bias, Individual Differences, And Normality*, 94 *ORGAN. BEHAV. HUM. DECISION PROCESS* 74 (2004).

²⁵⁹ Christopher J. Anderson, *The Psychology of Doing Nothing: Forms of Decision Avoidance Result from Reason and Emotion*, 129 *PSYCHOL. BULL.* 139 (2003).

²⁶⁰ Jonathan Breslin, *The Status Quo Bias And Decisions To Withdraw Life-Sustaining Treatment*, 190 *CMAJ E265* (2018).

²⁶¹ Jonathan Baron & Ilana Ritov, *Reluctance To Vaccinate: Omission Bias And Ambiguity*, 3 *J. BEHAV. DECIS. MAKING* 263 (1990).

²⁶² Jeffrey J. Rachlinski et al., *Inside the Bankruptcy Judge’s Mind*, 86 *B.U. L.R.* 1227 (2006).

²⁶³ Robert A. Prentice & Jonathan J. Koehler, *A Normality Bias in Legal Decision Making*, 88 *CORNELL L. REV.* 583, 638 (2003).

²⁶⁴ Omri Ben-Shahar & John A. E. Pottow, *On the Stickiness of Default Rules*, 33 *FLA. ST. U. L. REV.* 651, 655 (2006) (noting that “[f]indings do, indeed, lend support to the conclusion that human beings are cognitively disposed to prefer a default legal rule in contractual negotiations, irrespective of the content of that legal rule.”); see also Jeffrey J. Rachlinski & Andrew J. Wistrich, *Gains, Losses, and Judges: Framing and the Judiciary*, 94 *NOTRE DAME L.R.* 521, 570 (2019) (noting that “judges “preference for the status quo goes well beyond a reasonable effort to avoid disrupting settled expectations.”).

may be admitted in criminal trials without any meaningful scientific validation because it has always been admitted—it is the legal default—and once established “the obstinacy of the status quo” can be difficult to overcome.²⁶⁵

The status quo bias may also help explain the reluctance of judges to consider new information—in this case the NAS and PCAST reports, as well as other scientific studies proffered by defendants suggesting that previously admitted forensic evidence has been shown to be unreliable or scientifically invalid. Precedent reflects the current state of affairs, and judges may often accept that state “because to do otherwise would require significant cognitive effort.”²⁶⁶ Justice Cardozo praised this aspect of precedent when he noted that [t]he labor of judges would be increased almost to the breaking point if every past decision could be reopened in every case, and one could not lay one’s own course of bricks on the secure foundation of the courses laid by others who had gone before him.”²⁶⁷ Like President Obama’s suits, “by relying on past decisions, judges can save significant time and effort and thereby consider far more cases than would otherwise be possible. Judges can turn to past analyses and avoid rethinking every aspect of a decision.”²⁶⁸

Indeed, a preference for the status quo is not always irrational as “the status quo may have achieved its status through superiority over alternatives.”²⁶⁹ Unfortunately, this cannot be said of microscopic hair analysis, whose use the NRC report found to be “highly unreliable” and of “limited probative value.”²⁷⁰ In a striking example of the unreliability of microscopic hair analysis, 17-year-old Santae Tribble was convicted of murder in 1980 largely based on FBI hair microscopy evidence. At Tribble’s trial, an FBI analyst testified that he had microscopically compared a hair found at the crime scene with Tribble’s hair and found a “match.”²⁷¹ As the prosecutor told the jury during closing arguments, there was “only one chance in 10 million” that the hair belonged to someone else.²⁷² Based almost entirely on this evidence, Tribble was found guilty and sentenced to 20 years to life; Tribble was paroled in 2003 after serving 23 years in prison. Nine years later, in 2012, he was exonerated after DNA testing proved that the 13 hairs found at the crime scene did not belong to Tribble—instead they belonged to three other people and a dog.²⁷³

After the DC Public Defender Service uncovered Tribble’s case, along with those of two other men, Kirk Odom, and Donald Gates, all of whom were wrongfully convicted based on FBI hair microscopy testimony, the Washington Post ran a series of articles detailing the flawed forensics conducted by FBI examiners.²⁷⁴ These efforts triggered a federal review and a few years later in

²⁶⁵ Raquel Fernandez & Dani Rodrik, *Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty*, 81 AM. ECON. REV. 1146, 1155 (1991).

²⁶⁶ Robert A. Prentice & Jonathan J. Koehler, *A Normality Bias in Legal Decision Making*, 88 CORNELL L. REV. 583, 639 (2003).

²⁶⁷ BENJAMIN N. CARDOZO, *THE NATURE OF THE JUDICIAL PROCESS* 149 (1921).

²⁶⁸ Oona A. Hathaway, *Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System*, 86 IOWA L. REV. 601, 626 (2001).

²⁶⁹ A. Moshinsky & Maya Bar-Hillel, *Loss Aversion And Status Quo Label Bias*, 28, SOCIAL COGNITION 191 (2010).

²⁷⁰ NRC REPORT at 161.

²⁷¹ Brandon L. Garrett, *Bad Hair: The Legal Response to Mass Forensic Errors*, 42 LITIGATION 32 (2016).

²⁷² Brandon L. Garrett, *Bad Hair: The Legal Response to Mass Forensic Errors*, 42 LITIGATION 32, 32 (2016).

²⁷³ Brandon L. Garrett, *Bad Hair: The Legal Response to Mass Forensic Errors*, 42 LITIGATION 32, 32 (2016).

²⁷⁴ Spencer S. Hsu, *Convicted Defendants Left Uninformed Of Forensic Flaws Found By Justice Dept.*, WASH. POST (April 16, 2012), https://www.washingtonpost.com/local/crime/convicted-defendants-left-uninformed-of-forensic-flaws-found-by-justice-dept/2012/04/16/gIQAWTcgMT_story.html.

2015, the Department of Justice and the FBI formally acknowledged that nearly every examiner in the FBI's microscopic hair comparison unit gave flawed testimony in at least 90% of trials in which they offered evidence against criminal defendants over two decades before 2000, when mitochondrial DNA testing on hair became routine at the FBI.²⁷⁵ Those cases included 32 defendants who had been sentenced to death, 14 of whom were executed or died in prison.²⁷⁶

Long before the FBI's remarkable admission, in 1995, Chief Judge Frank Seay was ahead of his time when he conducted a thorough analysis of microscopic hair comparison evidence in *Williamson v. Reynolds* and found that "[a]lthough the hair expert may have followed procedures accepted in the community of hair experts, the human hair comparison results in this case were, nonetheless, scientifically unreliable" under *Daubert*.²⁷⁷ Williamson's conviction for capital murder and his death sentence had already been affirmed on appeal and he was five days away from execution at the time of the court's decision.²⁷⁸ Williamson was subsequently exonerated by DNA evidence.²⁷⁹

In reaching its decision, the *Williamson* court observed "an apparent scarcity of scientific studies regarding the reliability of hair comparison testing," noting that the few available studies "tend to point to the method's *unreliability*."²⁸⁰ Furthermore, the court concluded that there were no existing standards for human hair identification and the technique depends upon the examiner's subjective opinion. The court reviewed multiple studies indicating a high error rate, as high as 67% in some samples, and noted that most police laboratories were incorrect on 4 out of 5 samples analyzed, an "accuracy level below chance." Finally, the court concluded that even the "general acceptance" standard could not be met, as "any general acceptance seems to be among hair experts who are generally technicians testifying for the prosecution, not scientists who can objectively evaluate such evidence."²⁸¹ More than 20 years later, the PCAST report similarly concluded that there is no "scientific basis for concluding that microscopic hair examination is a valid and reliable process."²⁸²

Despite this thorough and particularly scathing analysis of the reliability of microscopic hair comparisons, the *Williamson* opinion "was all but ignored by other courts."²⁸³ And notwithstanding the FBI's remarkable admission in 2015, it is still the case that "[t]he

²⁷⁵ Federal Bureau of Investigation, FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review (April 20, 2015), <https://www.fbi.gov/news/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review>; see also Spencer S. Hsu, FBI Admits Flaws in Hair Analysis over Decades, WASH. POST (Apr. 18, 2015), https://www.washingtonpost.com/local/crime/fbi-overstated-forensic-hair-matches-in-nearly-all-criminal-trials-for-decades/2015/04/18/39c8d8c6-e515-11e4-b510-962fcfab310_story.html.

²⁷⁶ Spencer S. Hsu, FBI Admits Flaws in Hair Analysis over Decades, WASH. POST (Apr. 18, 2015), https://www.washingtonpost.com/local/crime/fbi-overstated-forensic-hair-matches-in-nearly-all-criminal-trials-for-decades/2015/04/18/39c8d8c6-e515-11e4-b510-962fcfab310_story.html.

²⁷⁷ *Williamson v. Reynolds*, 904 F. Supp. 1529 (E.D. Okla. 1995).

²⁷⁸ *Williamson v. Reynolds*, 904 F. Supp. 1529 (E.D. Okla. 1995).

²⁷⁹ Paul C. Giannelli, *Microscopic Hair Comparisons: A Cautionary Tale*, 46 CRIM. L. BULL. 531 (2010).

²⁸⁰ *Williamson v. Reynolds*, 904 F. Supp. 1529, 1556 (E.D. Okla. 1995) (emphasis in original).

²⁸¹ *Williamson v. Reynolds*, 904 F. Supp. 1529, 1556 (E.D. Okla. 1995).

²⁸² PCAST REPORT at 120.

²⁸³ Paul C. Giannelli, *Forensic Science: Daubert's Failure*, 68 CASE W. RESV. L. REV. 869, 885 (2018).

overwhelming majority of courts have deemed such evidence admissible.”²⁸⁴ Moreover, as recently as 2022, microscopic hair analysis examiners argued that “it is still nonetheless reasonable to conclude that accurate and reliable comparative analysis of hair morphology is possible.”²⁸⁵ Indeed, despite the lack of research-informed decisions and the technique’s “sole reliance on personal experience rather than ... empirical data,” a recent study found that “the evidential value of hair evidence is still perceived highly by examiners.”²⁸⁶ This is despite a 2023 report by the National Exonerations finding that “[a]t least 129 people have been falsely convicted based at least in part on MHCA. This is a conservative figure.”²⁸⁷

Given the extensive evidence of the unreliability of microscopic hair comparison evidence, as well as its continued and widespread admissibility by most courts around the country, the Supreme Court of Kentucky may have been influenced by the status quo bias in *Meskimen v Kentucky*, when it held that the trial court did not abuse its discretion by declining to hold a *Daubert* hearing and taking judicial notice that microscopic hair comparison evidence is scientifically reliable.²⁸⁸ The court emphasized that courts should not “reinvent the wheel ... by requiring the parties to put on full demonstrations of the validity or invalidity of methods or techniques that have been scrutinized well enough in prior decisions to warrant taking judicial notice of their status.”²⁸⁹

In support of its holding, the *Meskimen* court cites to a 1999 Kentucky Supreme Court decision, noting that “[e]vidence of hair analysis by microscopic comparison has been admissible in this Commonwealth for many years.”²⁹⁰ In that case, *Johnson v Commonwealth*, the Kentucky Supreme Court also observed that “[e]vidence of hair analysis by microscopic comparison has been admissible in this Commonwealth for many years,” and cited to cases dating back to 1950 upholding the admissibility of this type of evidence.²⁹¹ Notably, the *Johnson* court itself concedes that it has “never specifically addressed the scientific reliability of this method of hair analysis...[but] we must assume that it at least satisfied the *Frye* test of general acceptance; for otherwise, the evidence would never have been admitted in the first place.”²⁹²

In an impressive example of circular reasoning, the *Johnson* court goes on to explain that “[t]he absence in our previous opinions of any in-depth analysis under the “general acceptance” test was probably due to the overwhelming acceptance of this procedure as a reliable scientific method for the past fifty years.”²⁹³ The court then cites to a 1983 ALR and a string of cases from other jurisdictions upholding microscopic hair analysis evidence to support its conclusion that “trial

²⁸⁴ *State v. West*, 274 Conn. 605, 635 (2005).

²⁸⁵ Emma Redman et al., *An Objective and Statistical Approach to Microscopic Human Hair Comparison: A Laboratory Exercise for the Forensic Science Undergraduate and Graduate Student*, 4 J. FORENSIC SCI. EDUC. 1, 2 (2022).

²⁸⁶ Laura Wilkinson & Claire Gwinnett, *An International Survey Into The Analysis And Interpretation Of Microscopic Hair Evidence By Forensic Hair Examiners*, 308 FORENSIC SCI. INT’L. 1, 14 (2020).

²⁸⁷ NATIONAL REGISTRY OF EXONERATIONS, MICROSCOPIC HAIR COMPARISON ANALYSIS AND CONVICTING THE INNOCENT at 7 (2023), <https://www.law.umich.edu/special/exoneration/Documents/NREReportMHCA.pdf>

²⁸⁸ *Meskimen v. Com.*, 435 S.W.3d 526 (Ky. 2013).

²⁸⁹ *Meskimen v. Com.*, 435 S.W.3d 526, 535 (Ky. 2013).

²⁹⁰ *Johnson v. Com.*, 12 S.W.3d 258 (1999).

²⁹¹ E.g., *Wilhite v. Com.*, 574 S.W.2d 304 (Ky. 1978); *Sherley v. Com.*, 558 S.W.2d 615 (1977); *Garr v. Com.*, 463 S.W.2d 109 (1971); *Mitchell v. Com.*, 280 S.W.2d 189 (1955); *Acrey v. Com.*, 229 S.W.2d 748 (1950).

²⁹² *Johnson v. Com.*, 12 S.W.3d 258, 262 (Ky. 1999).

²⁹³ *Johnson v. Com.*, 12 S.W.3d 258, 262 (Ky. 1999).

courts in Kentucky can take judicial notice that this particular method or technique is deemed scientifically reliable.” The *Johnson* court does not mention either the NRC report, or the Oklahoma district court’s decision in *Williamson*, both of which were publicly available to the court at the time of its decision.

Although *Daubert* is meant to be a flexible test, this flexibility is not “a license to scrutinize sloppily or not at all.”²⁹⁴ And while the *Meskimen* court may not have known of the *Williamson* decision or of specific research findings suggesting the scientific unreliability of the forensic technique, by the time of the appellate court’s decision, Tribble, Odom, and Gates had been exonerated and the Washington Post’s extensive coverage of the three cases had caused the FBI to begin a major investigation into the reliability of microscopic hair evidence. As the *Meskimen* court itself noted, even when case law supports the admissibility of a forensic method, “it is the trial court’s duty to ensure that the method is supported by scientific findings, or at least not seriously questioned by recent reputable scientific findings” before taking judicial notice of the technique.²⁹⁵ By declining to even hold a *Daubert* hearing, the trial court could not possibly have met this obligation, yet the *Meskimen* court, without itself attempting to consider any evidence to the contrary, found that the trial court’s decision was not an error.

Although the status quo bias can make decision-making more efficient by preventing the need to “reinvent the wheel,” it can also be harmful if “a non-optimal current situation continues or if helpful improvements are ignored.”²⁹⁶ In *Meskimen*, the judge may have been disinclined to “reinvent the wheel” in order to avoid the significant cognitive effort associated with reviewing the scientific validity of a forensic technique that has been admitted “for the past 50 years” and is still widely perceived to be reliable by forensic examiners. Similarly, because the admission of microscopic hair evidence in Kentucky is the legal default, one that the judge in *Meskimen* valued and understood, they may have been less likely to opt out of that default and instead stick with the status quo. In either case, the *Meskimen* court’s preference for the status quo—in this case the continued admissibility of microscopic hair evidence—is apparent, even though a change could lead to better outcomes, namely the avoidance of dozens of wrongful convictions based on flawed forensic science.

C. Forensic Firearms Analysis & the Omission Bias

While status quo bias manifests as a preference for the current situation and an aversion to change, the closely related omission bias is driven by the fear of regret and responsibility for one’s actions.²⁹⁷ In this way people prefer inaction over action because they anticipate experiencing more regret if they act and the outcome is poor than if their failure to act results in harm.

We can see examples of our preference for inaction and the omissions bias in a variety of domains. Most legal systems do not hold people liable for a failure to rescue or intervene absent a legal duty

²⁹⁴ Michael J. Saks & David L. Faigman, *Expert Evidence After Daubert*, 1 AN. REV. L. & SOCIAL SCI. 105, 108 (2005).

²⁹⁵ *Meskimen v. Com.*, 435 S.W.3d 526, 535 (Ky. 2013).

²⁹⁶ Marie E. Godefroid et al., *How To Measure The Status Quo Bias? A Review Of Current Literature*, 73 MANAG. REV. Q. 1667, 1668 (2023).

²⁹⁷ Maurice Schweitzer, *Disentangling Status Quo and Omission Effects: An Experimental Analysis*, 58 ORGANIZATIONAL BEHAV. HUM. DECISION PROC. 457 (1994).

to act.²⁹⁸ And many people consider it worse to lie by commission by stating something false than to lie by omission by withholding important yet true information.²⁹⁹ Parents who chose not to vaccinate their children for whooping cough are more likely than others to think that vaccinating was more dangerous than not vaccinating, even though the chance of harm was greater than if they were to vaccinate.³⁰⁰ Sports referees also exhibit a preference for inaction when making calls. As one NBA referee explained, “if it’s late in the game and, let’s say, there’s goaltending and you miss it. That’s an incorrect call and that’s bad. But let’s say it’s late in the game and you call goaltending on a play and the replay shows it was an incorrect call. That’s when you’re in a *really* deep mess.”³⁰¹

When influenced by the omission bias, people do not act because they believe they are more responsible for their actions than for their inactions, and the regret they anticipate feeling from action is therefore higher than from inaction.³⁰² In this way, negative outcomes will be seen as worse when they result from action rather than omission, and positive outcomes will be more gratifying when they result from action rather than omission. “Faced with uncertainty, people who are loss averse thus have a bias towards omissions.”³⁰³ High levels of anticipated regret can motivate individuals “to seek escape by means of a justification such as the default option or by prolonging the decision in hopes of avoiding responsibility for it.”³⁰⁴ Moreover, when coping with the stress generated by a difficult decision, decisionmakers sometimes exhibit buck passing, in which responsibility for the decision is shifted to others.³⁰⁵

Research in legal decision-making suggests that factfinders exhibit an omission bias, which may it more likely that a judge will affirm a prior ruling or follow existing precedent, even if they may have decided differently had they been the first to try the case. For example, while the general standard of proof in civil litigation is preponderance of the evidence, Zamir and Ritov found that that participants in their studies employed a standard of proof that was considerably higher. They attributed this discrepancy to omission bias, “since ruling in favor of the plaintiff is framed as actively interfering in the usual course of things and dismissing a claim is seen as refraining from action, the omission bias inhibits the judge from ruling in favor of the plaintiff unless her case is truly compelling.”³⁰⁶ Similarly, other researchers suggest that given the omission bias, “it seems

²⁹⁸ See e.g. *Twitter, Inc. v. Taamneh*, 598 U.S. 471, 489 (2023) (noting that “our legal system generally does not impose liability for mere omissions, inactions, or nonfeasance; although inaction can be culpable in the face of some independent duty to act, the law does not impose a generalized duty to rescue.”).

²⁹⁹ See Mark Spranca et al., *Omission and Commission in Judgment and Choice*, 27 J. EXPERIMENTAL SOC. PSYCHOL. 76, 102 (1991); Lyn M. van Swol et al., *Effects of Gains/Loss Frames on Telling Lies of Omission and Commission*, 36 COGNITION & EMOTION 1287, 1296 (2022).

³⁰⁰ David A. Asch et al., *Omission Bias and Pertussis Vaccination*, 14 MED. DECIS. MAKING 118, 121 (1994).

³⁰¹ TOBY MOSKOWITZ AND L. JOHN WERTHEIM, *SCORECASTING: THE HIDDEN INFLUENCES BEHIND HOW SPORTS ARE PLAYED AND GAMES ARE WON* 11 (Random House 2011).

³⁰² David A. Asch et al., *Omission Bias and Pertussis Vaccination*, 14 MED. DECIS. MAKING 118, 121 (1994).

³⁰³ Eval Zamir & Ilana Ritov, *Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation*, 41 J. LEGAL STUDIES 165, 165 (2012).

³⁰⁴ Christopher Anderson, *The Psychology of Doing Nothing: Forms of Decision Avoidance Result From Reason and Emotion*, 129 PSYCHOL. BULL 139, 148 (2003).

³⁰⁵ Christopher Anderson, *The Psychology of Doing Nothing: Forms of Decision Avoidance Result From Reason and Emotion*, 129 PSYCHOL. BULL 139, 140 (2003).

³⁰⁶ Eval Zamir & Ilana Ritov, *Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation*, 41 J. LEGAL STUDIES 165, 166 (2012).

reasonable to predict that judges would favor affirmances over reversals.³⁰⁷ Leaving a lower court opinion undisturbed is akin to an omission, which reversing the decision is akin to a commission. The omission bias—in conjunction with the status quo bias—can therefore cause judges to favor affirmances over reversals, thus “protecting themselves psychologically because any negative outcome resulting from the decision is a product of the trial judge’s action, not the appellate judge’s inaction.”³⁰⁸

The omission bias may have been at play in *United States v Green*, a case in which the government sought to introduce ballistics testimony from Detective O’Shea of the Boston Police Department. O’Shea was prepared to testify that the shell casings found at a crime scene were an exact match to a weapon found in front of the defendant’s home, and that this match could be made “to the exclusion of every other firearm in the world.”³⁰⁹ The trial court judge in *Green* was clearly uneasy with O’Shea’s proposed testimony, noting that O’Shea’s conclusion was “extraordinary, particularly given [his] data and methods.”³¹⁰

Following a *Daubert* hearing, the judge concluded that while O’Shea had seven years of experience in the Boston Police Ballistics unit, he had never been certified by any professional organization. Detective O’Shea could not provide the court with any information about error rates, either from his own laboratory, or from the field in general. The judge also questioned the validity of toolmark evidence itself, noting that the premise of the field—that the surface contours of each firearm are unique—is itself contested, and that in any event, the “task of telling [shell casings] apart is not an easy one” and is largely subjective; in this case, Detective O’Shea “relied mainly on his subjective judgment.” Moreover, O’Shea conducted what the court described as an evidentiary “show-up,” comparing the weapon police suspected belonged to the Defendant to shell casings found at a crime scene, “not what scientists would regard as a “blind” test.”³¹¹

Following this extensive review of the “serious deficiencies” of the proposed evidence, the court nevertheless concluded that “*every single court post-Daubert* has admitted this testimony” and that given this precedent, and “and notwithstanding my serious reservations, I feel compelled to allow O’Shea to testify about his observations of the shell casings.”³¹² Finally, the judge explained that he “reluctantly c[a]me to the above conclusion because of my confidence that any other decision will be rejected by appellate courts, in light of precedents across the country, regardless of the findings I have made.”³¹³

³⁰⁷ Chris Guthrie & Tracey E. George, *Judicial Decisionmaking: The Futility of Appeal: Disciplinary Insights into the “Affirmance Effect” on the United States Courts of Appeals*, 32 FL. ST. UNIV. L.R. 357 (2005).

³⁰⁸ Chris Guthrie & Tracey E. George, *Judicial Decisionmaking: The Futility of Appeal: Disciplinary Insights into the “Affirmance Effect” on the United States Courts of Appeals*, 32 FL. ST. UNIV. L.R. 357, 380 (2005).

³⁰⁹ *U.S. v. Green*, 405 F. Supp. 2d 104, 107 (D. Mass. 2005).

³¹⁰ *U.S. v. Green*, 405 F. Supp. 2d 104, 107 (D. Mass. 2005).

³¹¹ *U.S. v. Green*, 405 F. Supp. 2d 104, 108 (D. Mass. 2005).

³¹² *U.S. v. Green*, 405 F. Supp. 2d 104, 108 (D. Mass. 2005) (emphasis in original).

³¹³ Somewhat surprisingly, the judge then admonishes other courts to “require more,” noting that “[t]he more courts admit this type of toolmark evidence without requiring documentation, proficiency testing, or evidence of reliability, the more sloppy practices will endure.” *Id.* at 109.

Lower courts judges may be individually motivated to follow precedent out of concern for their own reputations or because they fear being overruled.³¹⁴ However, given that *Daubert* and 702 permit the exclusion of evidence if the evidence is unreliable in the case at issue, regardless of prior decisions, and *Joiner* requires appellate courts to use an abuse of discretion standard when reviewing trial court rulings on admissibility under *Daubert*, this fear is probably largely unfounded.

Moreover, it may also be the case that the omission bias played a role in the judge's decision in *Green*. The possible "rejection" of his decision by an appellate court is a negative outcome the judge may consider worse because it would result from action rather than omission and he therefore chooses not to act. Or like the decisionmakers in Zamir and Ritov's study, because "the usual course of things" is the admission of ballistics evidence, the omission bias may have inhibited the judge from ruling for Green, despite the judge's serious reservations. Similarly, a high level of anticipated regret, given his confidence that his decision would be rejected by an appellate court, may have caused the judge to stick with the default option of admitting the government's ballistics evidence in hopes of avoiding responsibility for making the decision to exclude the evidence.

Finally, courts have been admitting ballistics evidence for nearly 200 years. The first known toolmark comparison case occurred in 1835 in London, when police investigating a shooting identified a flaw in the fired bullet and compared it to the gun manufacturer's mold.³¹⁵ In the United States, expert testimony regarding firearms identifications has been admissible for over a century.³¹⁶ And as discussed above, while both the 2009 NRC report and the 2016 PCAST report have expressed significant doubt about the validity of ballistics matching, the court in *Green* was correct that the majority of jurisdictions continue to allow the admission of toolmark comparison evidence without restrictions or limitations on the form of the expert testimony.³¹⁷

It is therefore not entirely surprising that the judge in *Green*, despite his serious concerns about the reliability of the evidence at hand, was reluctant to act. To exclude toolmark comparison evidence against that historical backdrop would require the judge to acknowledge that this type of evidence has been improperly admitted for over a century in hundreds of cases. And perhaps just as importantly, excluding this type of evidence would require action, and the omissions bias causes us to view actions as worse than omissions, regardless of the consequences of the act or the omission. The judge may have decided the case differently had he been the first to consider the scientific validity of the proffered ballistics evidence, but instead seems to have viewed that harm that could result from finding the evidence inadmissible (being overruled) more negatively than the harm that could result from doing nothing (the admission of unreliable evidence), and therefore chose not to act.

³¹⁴ Matthew Tokson, *Judicial Resistance and Legal Change*, 82 U. CHICAGO L. REV. 901, 910 (2015).

³¹⁵ Firearms Examiner Training, https://projects.nfstc.org/firearms/module02/fir_m02_t04.htm (last visited Aug. 1, 2024).

³¹⁶ See e.g., *Commonwealth v. Best*, 62 N.E. 748, 750 (Mass. 1902); *State v. Clark*, 196 P. 360, 367–69 (Or. 1921); *Laney v. United States*, 294 F. 412, 416 (D.C. Cir. 1923).

³¹⁷ Jim Agar, *The Admissibility Of Firearms And Toolmarks Expert Testimony In The Shadow Of PCAST*, 74 BAYLOR L.R. 93, 143-44 (2022). Some courts have begun restricting analysts from using the word "match" or claiming that the bullet in question was fired by a specific gun to the exclusion "of all other guns in the world." See e.g., *United States v. Ashburn*, 88 F. Supp. 3d 239, 247 (E.D.N.Y. 2015).

III. HOW CAN WE RESPOND TO THE CONTINUED ADMISSION OF UNRELIABLE FORENSIC EVIDENCE?

Legal scholars and legislators have suggested a variety of reforms to address the continued admissibility of unreliable forensic evidence. Some have expressed support for the “junk science writ,” laws that would allow courts to review post-conviction relief petitions based on faulty science that was later discredited.³¹⁸ One prominent proponent of such a remedy is Justice Sonya Sotomayor, who released a statement in the 2024 case of *Alabama v. McCrory* denying review of a conviction based almost exclusively on bite-mark testimony. Although Justice Sotomayor joined her colleagues in denying McCrory’s due process claim, she noted that,

Legislatures concerned with wrongful convictions based on faulty science, however, need not wait for this Court to address a constitutional remedy. Several States have already tackled this troubling problem through targeted postconviction statutes. These statutes create an efficient avenue for innocent people convicted based on forensic science that the scientific community has now largely repudiated.”³¹⁹

State legislatures in Texas³²⁰ and California³²¹ have both passed such a law. Unfortunately, a 2024 report from the Texas Defender Service (TDS) examined more than 70 cases raised under the Texas statute between September 2013 and December 2023 and found that “[a]rticle 11.073 is not working to provide relief to innocent people convicted based on false or unreliable forensic evidence.”³²² In particular, the TDS Report concluded that the Texas Court of Criminal Appeals implementation of the report has shown “a disregard for discredited scientific methods,” and that individuals seeking relief under the statute “might need to go beyond proving the State’s reliance on flawed science—they might need to provide evidence affirmatively showing innocence.”³²³ Of the 73 cases the report reviewed, just 15 applicants received relief from the courts and in 73% of those cases, the person seeking relief “affirmatively prove[d] their actual innocence.”³²⁴ Moreover, the report found that the Court of Appeals largely restricts relief to those cases involving new DNA evidence, despite the fact that most wrongful conviction cases do not involve DNA.³²⁵

³¹⁸ See e.g., Valena E. Beety, *Changed Science Writs and State Habeas Relief*, 57 HOUS. L. REV. 483 (2020).

³¹⁹ *McCrory v. Alabama*, 144 S. Ct. 2483 (2024); see also <https://theintercept.com/2024/07/10/bite-mark-supreme-court-sotomayor/>, Liliana Segura & Jordan Smith, *There’s a Junk Science Crisis in Criminal Convictions. Sonia Sotomayor Calls it out in Alabama Bite-Mark Case*, *The Intercept* (July 10, 2024), <https://theintercept.com/2024/07/10/bite-mark-supreme-court-sotomayor/>.

³²⁰ Tex. Code Crim. Proc. art. 11.073 (“(a) This article applies to relevant scientific evidence that: (1) was not available to be offered by a convicted person at the convicted person’s trial; or (2) contradicts scientific evidence relied on by the state at trial.”).

³²¹ Cal. Penal Code § 1473 (D)(ii) (“A significant dispute has emerged or further developed in the petitioner’s favor regarding expert medical, scientific, or forensic testimony that was introduced at trial or a hearing and that expert testimony more likely than not affected the outcome of the case. (ii) For purposes of this section, the significant dispute may be as to the reliability or validity of the diagnosis, technique, methods, theories, research, or studies upon which a medical, scientific, or forensic expert based their testimony.”).

³²² TEXAS DEFENDER SERVICE, AN UNFULFILLED PROMISE: ASSESSING THE EFFICACY OF 11.073 at 1 (July 2024), available at <https://www.texasdefender.org/wp-content/uploads/2024/07/TDS-11.073-Report.pdf> [hereinafter TDS REPORT].

³²³ TDS REPORT at 12.

³²⁴ TDS REPORT at 11.

³²⁵ The Nat’l Registry of Exonerations, *Exonerations by Year, DNA and non-DNA*, <https://www.law.umich.edu/special/exoneration/Pages/Exoneration-by-Year.aspx> (last visited Dec. 11, 2024).

As an example of a case in which evidence of flawed science has not been enough to secure relief under the Texas statute, the TDS Report describes the case of Robert Roberson, who has spent more than 20 years on death row in Texas for killing his two-year-old daughter Nikki. Roberson was convicted in 2003 based on expert testimony that Nikki died from “shaken baby syndrome,” a diagnosis about which “the scientific knowledge has evolved...based on twenty years of reputable scientific studies and publications.”³²⁶ Although Roberson sought relief under Article 11.073 multiple times, the Texas Criminal Court of Appeal—which has itself acknowledged that the scientific basis for shaken baby syndrome has been called into significant question³²⁷—has continually denied relief.³²⁸ In support of Roberson’s request to the Board of Pardons and Paroles for clemency, 86 members of the Texas House of Representatives wrote to explain that the House unanimously passed Article 11.073 to allow “challenges to convictions that were based on disproven or incomplete science,” and that they were “dismayed to learn that this law has not been applied as intended and has not been a pathway to relief—or even a new trial—for people like Mr. Roberson.”³²⁹

Although there is no clear indication that cognitive bias played a role in the outcome in Roberson’s case, as the Texas Representatives’ letter notes, “it appears that both the trial court and the Court of Criminal Appeals refused to acknowledge or engage with this voluminous new scientific evidence and instead simply denied Mr. Roberson a new trial.”³³⁰ Roberson’s case—and the failure of the Texas law to meaningfully address the admission of flawed forensic evidence—highlights the need to go beyond these sorts of legislative reforms and simultaneously address the admission of unreliable forensic evidence at the judicial level.

On December 1, 2023, Rule 702 itself was also amended to address what the Advisory Committee on Evidence viewed as a “widespread misunderstanding in the case law.”³³¹ Specifically, Rule 702 now requires that the proponent of expert testimony “demonstrates to the court that it is more likely than not that” each of the requirements of Rule 702 have been met.³³² In addition, 702(d) now

³²⁶ *Ex parte Roark*, No. WR-56,380-03, 2024 WL 4446858, at *26 (Tex. Crim. App. Oct. 9, 2024).

³²⁷ *Ex parte Roark*, No. WR-56,380-03, 2024 WL 4446858, at *26 (Tex. Crim. App. Oct. 9, 2024) (noting that were the experts in Roark’s case to testify today as they did at his original trial, they would be “confronted with twenty years of reputable scientific studies and publications that . . . contradicts their trial testimony.”).

³²⁸ For an excellent overview of the complicated history of the Roberson case, see the statement of Justice Sonya Sotomayor in *Roberson v Texas*, No. 24-5753, 2024 WL 4521766, at *1 (U.S. Oct. 17, 2024).

³²⁹ See Letter from 86 Members of the Texas House of Representatives to the Texas Board of Pardons & Paroles (Sept. 16, 2024), available at <https://drive.google.com/file/d/1UvvlNn7TaNptvZPuyStJRtMqWH-ftY9n/view> (last visited Dec. 10, 2024). Roberson was scheduled to be executed on October 17, 2024, but on October 16, the Texas House Committee on Criminal Jurisprudence issued a subpoena for him to testify the day after his scheduled execution and the district court temporarily halted the execution so Roberson could testify. As of December 2024, the State of Texas has not yet set a new execution date. Michael Hall, Texas Legislators Kept Robert Roberson—and Bipartisanship—Alive (for Now), *Texas Monthly* (Jan. 2025), available at <https://www.texasmonthly.com/news-politics/robert-roberon-bipartisan-execution-delay-effort/>.

³³⁰ Letter from 86 Members of the Texas House of Representatives to the Texas Board of Pardons & Paroles (Sept. 16, 2024), available at <https://drive.google.com/file/d/1UvvlNn7TaNptvZPuyStJRtMqWH-ftY9n/view> (last visited Dec. 10, 2024).

³³¹ Advisory Comm. on Evidence Rules, *Agenda for Committee Meeting* 7 (Apr. 30, 2021); see also Fed. R. Evid. 702.

³³² As the Advisory Committee explained, courts often find expert testimony admissible even though the proponent has not satisfied 702’s requirements by a preponderance of the evidence—“essentially treating these questions as ones of weight rather than admissibility,” contrary to Rule 104(a). Committee on Rules of Practice and Procedure 871 (June 7, 2022), https://www.uscourts.gov/sites/default/files/2022-06_standing_committee_agenda_book_final.pdf.

states that the expert’s opinion must reflect “a reliable application” of principles and methods. The Advisory Committee proposed these amendment to stress the gatekeeping function that judges play under Rule 702 and to ensure that courts perform a Rule 702 analysis before admitting an expert opinion. As the new rule attempts to clarify, courts cannot simply refer to Rule 702 and then admit expert testimony without finding by preponderance that the evidence meets all the requirements of the rule.³³³

The Fourth Circuit Court of Appeals was an early adopter of the 2023 amendments to Rule 702, holding in a 2021 wrongful death case that the trial court had “improperly abdicated its critical gatekeeping role to the jury and admitted [the testimony of plaintiffs’ experts] without engaging in the required Rule 702 analysis.”³³⁴ Citing to the pending amendments to Rule 702, the court confirmed “once again the indispensable nature of district courts’ Rule 702 gatekeeping function in all cases in which expert testimony is challenged on relevance and/or reliability grounds.”³³⁵ As commentators noted at the time, the 2023 amendments are important because they “effectively overrule a significant body of case law that has misconstrued and misapplied Rule 702 for more than 20 years.”³³⁶ In particular, those cases that fail to appropriately apply Rule 702 to the admission of forensic evidence “should no longer be cited.”³³⁷

But while legislative reforms like the junk science writ show may eventually help to remedy past convictions based on discredited forensic evidence, and the 2023 Amendments may inspire courts like the Fourth Circuit to begin dismantling the flawed trail of precedent admitting that evidence going forward, other steps should also be taken to address the impact of cognitive bias on the admission of flawed forensic evidence and better fulfill the mandates of Rule 702. First, this Article argues that increasing diversity on the bench is a critical reform necessary to better fulfill the requirements of Rule 702. Greater representation of underrepresented groups in the judiciary can introduce varied perspectives, challenging entrenched biases and fostering a more critical approach to long-standing forensic practices. Next, judicial training programs should be enhanced to include more education on scientific standards. By providing judges with the tools to better understand and interpret the scientific principles and methodology underlying various forensic methods, judges will be better equipped to evaluate the reliability and ultimate admissibility of that evidence. In addition to better scientific education, judges should also be better educated on the ways in which cognitive biases and heuristics influence their decision-making. Although awareness of bias is generally insufficient to remove bias from decision-making, it can be the first step toward debiasing, and judges should also be trained in specific debiasing strategies to help mitigate the effect of cognitive bias on their decision-making.

A. Increased Diversity on the Bench

³³³ The Advisory Committee was also concerned about the tendency of experts to overstate their conclusions and amended Rule 702(d) to “emphasize that each expert opinion must stay within the bounds of what can be concluded from a reliable application of the expert’s basis and methodology.” Specifically, the commentary states that “[f]orensic experts should avoid assertions of absolute or one hundred percent certainty—or to a reasonable degree of scientific certainty—if the methodology is subjective and thus potentially subject to error.” Fed. R. Evid. 702.

³³⁴ *Sardis v. Overhead Door Corp.*, 10 F.4th 268, 282 (4th Cir. 2021).

³³⁵ *Sardis v. Overhead Door Corp.*, 10 F.4th 268, 282 (4th Cir. 2021).

³³⁶ Eric Lasker & Lawrence Ebner, *It’s Time for Attorneys and Courts to use the Amended Expert Witness Rule* (July 22, 2023), <https://atlanticlegal.org/news/its-time-for-attorneys-courts-to-use-the-amended-expert-witness-rule/>.

³³⁷ *Id.*

According to the 2023 ABA Profile of the Legal Profession, the federal bench in the United States is still largely white and male. Although President Biden has nominated a much more diverse group on federal judges since 2021, 68% of all sitting Article III judges are male and 76% are white. Only 4.7% of all federal judges are Black women and just four judges identify as Native American.³³⁸ State court judicial demographics reflect similar disparities. Among state supreme court justices in 2022, in 20 states there are no justices who identify as a person of color, including in 12 states where people of color constitute at least 20% of the state population. Men still hold 59% of state supreme court seats, and across all state high courts, only 18% of all justices are Latino, Asian American, Native American, or multiracial.³³⁹ With respect to professional diversity, 39% of sitting state supreme court justices are former prosecutors, while only 7% are former public defenders.³⁴⁰

The benefits of a diverse judiciary are extensive and well documented.³⁴¹ A diverse judiciary incorporates a broader range of experience and perspectives into the interpretation and application of the law. By virtue of their outsider status, women, visible minorities, and other historically marginalized groups are able to recognize the assumptions and stereotypes steeped in legal doctrine and precedent that others may not see.³⁴² The same can be said of a judiciary that reflects geographic, socioeconomic, and political diversity. Moreover, a judiciary that reflects the demographic characteristics of its population serves a legitimizing function and increases public confidence in the courts.³⁴³

In addition to these benefits, increasing diversity in the judiciary may help to dismantle unreliable precedent as people from diverse backgrounds can bring a different perspective to precedent that others might not see the need to question.³⁴⁴ As Sherrilyn Ifill once observed, “the creation of a racially diverse bench can introduce traditionally excluded perspectives and values into judicial decision-making. The interplay of diverse views and perspectives can enrich judicial decision-making.”³⁴⁵ Moreover, diverse court “is a smart court - one that is more likely to be innovative, productive and efficient in meeting the challenges facing the justice system in the twenty-first century.”³⁴⁶ Although it is not clear that characteristics like race and gender systematically affect

³³⁸ ABA Profile of the Legal Profession 2023, <https://www.abalegalprofile.com/judges.html> (last visited Aug. 1, 2024).

³³⁹ Brennan Center for Justice, State Supreme Court Diversity—May 2022 Update, <https://www.brennancenter.org/our-work/research-reports/state-supreme-court-diversity-may-2022-update>.

³⁴⁰ *Id.*

³⁴¹ See e.g. Kevin Johnson & Luis Fuentes-Rohwer, *A Principled Approach to the Quest for Racial Diversity on the Judiciary*, 10 MICH. J. RACE & LAW 5, 24 (2004); Honorable Edward M. Chen, *The Judiciary, Diversity, and Justice for All*, 10 ASIAN L.J. 127 (2003); Lady Hale, *Making a Difference? Why We Need a More Diverse Judiciary*, 56 N. IR. LEGAL Q. 281 (2003); K.O. Myers, *Merit Selection and Diversity on the Bench*, 46 IND. L. REV. 43, 45-46 (2013) (summarizing a “growing body of research looking at the benefits of a diverse judiciary.”).

³⁴² Mark S. Hurwitz & Drew Noble Lanier, *Diversity in State and Federal Appellate Courts: Change and Continuity Across 20 Years*, 29 JUSTICE SYSTEM J. 47, 49 (2008) (noting that “there may be systematic, empirically discernible differences in the decision making of judges of nontraditional back-grounds.”).

³⁴³ Mark S. Hurwitz & Drew Noble Lanier, *Diversity in State and Federal Appellate Courts: Change and Continuity Across 20 Years*, 29 JUSTICE SYSTEM J. 47, 49 (2008).

³⁴⁴ Erin P. Hennes & Layla Dang, *The Devil We Know: Legal Precedent and the Preservation of Injustice*, 8 POLICY INSIGHTS BEHAVIORAL BRAIN SCI. 76, 80 (2021);

³⁴⁵ Sherrilyn A. Ifill, *Racial Diversity on the Bench: Beyond Role Models and Public Confidence*, 57 WASHINGTON & LEE L.R. 405, 410 (2000).

³⁴⁶ SHERYL J. WILLERT, BUILDING A DIVERSE COURT: A GUIDE TO RECRUITMENT AND RETENTION *at v* (June 2010).

a judge's decision-making, many authors argue that diversity improves decision-making across a variety of domains.³⁴⁷ Moreover, as Jeffrey J. Rachlinskit has observed, “[g]iven the complexity of human cognition and the incredible variation in attitudes, beliefs, intelligence, and experience,” different decisionmakers are almost certainly not committing identical errors.”³⁴⁸

Finally, the benefits of diversity on group decision-making are well-documented. For example, many studies highlight the benefits of diverse juries, which deliberate longer and consider more information than juries that are less diverse.³⁴⁹ The same has been found to be true in the workplace,³⁵⁰ in the armed forces,³⁵¹ and in healthcare.³⁵² Diversity also improves creativity within the group, flexibility in decision-making, and thoughtfulness.³⁵³ While individuals from diverse groups are not immune to cognitive biases, they may be more willing to revisit institutional barriers to change, which in turn could lead to a reduction in bias and better decision-making.³⁵⁴

While many decisions about the admission of forensic evidence are made by individual judges, judges often justify their decisions in written opinions, which “forces them to engage with precedent and refine their own views.”³⁵⁵ In this way, individual judges are part of a larger discourse in which they are influenced by their peers. When those peers are able to draw from a broader range of knowledge, expertise, and perspectives, this in turn may lead to higher quality decision-making and more careful consideration of the quality of earlier decisions about the

³⁴⁷ See, e.g. SCOTT PAGE, *THE DIFFERENCE: HOW THE POWER OF DIVERSITY CREATES BETTER GROUPS, FIRMS, SCHOOLS, AND SOCIETIES* (Princeton Univ. Press 2007); Mark S. Hurwitz & Drew Noble Lanier, *Diversity in State and Federal Appellate Courts: Change and Continuity Across 20 Years*, 29 JUSTICE SYSTEM J. 47, 49 (2008) (noting that “there may be systematic, empirically discernible differences in the decision making of judges of nontraditional backgrounds.”).

³⁴⁸ Jeffrey J. Rachlinskit, *Cognitive Errors, Individual Differences, and Paternalism*, 73 UNIV. CHICAGO L.R. 207, 208 (2006). Or as Justice Benjamin Cardozo once noted, “the eccentricities of judges balance one another.” BENJAMIN CARDOZO, *THE NATURE OF THE JUDICIAL PROCESS* 177 (Yale Univ. Press 1921).

³⁴⁹ Samuel R. Sommers, *On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations*, 90 J. PERSONALITY & SOC. PSYCHOL. 597, 606 (2006).

³⁵⁰ See, e.g., Elizabeth Mannix & Margaret A. Neale, *What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations*, 6 PSYCHOL. SCI. PUB. INT. 31, 34 (2005) (reviewing the literature and concluding that deep-level diversity has “typically been shown to improve performance through vigorous debate that leads to creativity and improved problem solving”).

³⁵¹ Tara A. (Rench) Brown et al., *Contextualizing Inclusion: Developing a Framework and Measure for a Military Context*, 32 MILITARY PSYCHOL. 313 (2020).

³⁵² See e.g., L. E. Gomez & Patrick Bernet, *Diversity Improves Performance and Outcomes*, 111 J. NAT’L MED. ASSOC. 383, 391 (2019) (reviewing studies focused on the impact of diversity in the workplace and concluding that “diversity improves performance. It is associated with higher profits and a range of financial rewards including: innovation, increased productivity, improved accuracy in risk assessment and has already been associated with improved patient health outcomes.”).

³⁵³ Samuel R. Sommers, *On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations*, 90 J. PERSONALITY & SOC. PSYCHOL. 597, 598 (2006) (citing various studies); Bernard A. Nijstad & Silvia C. Kaps, *Taking the Easy Way Out: Preference Diversity, Decision Strategies, and Decision Refusal in Groups*, 94 J. PERSONALITY & SOC. PSYCHOL. 860, 860 (2008) (noting that “groups have the potential to outperform individuals when it comes to the quality of decisions”).

³⁵⁴ Erin P. Hennes & Layla Dang, *The Devil We Know: Legal Precedent and the Preservation of Injustice*, 8 POLICY INSIGHTS BEHAVIORAL BRAIN SCI. 76, 80 (2021).

³⁵⁵ Jason Iuliano & Avery Stewart, *The New Diversity Crisis in the Federal Judiciary*, 84 TENN. L.R. 247, 265 (2016).

admissibility of forensic evidence for all judges.³⁵⁶ Moreover, when judges on appellate panels or state or federal supreme courts review trial court decisions on the admissibility of forensic evidence, judges deliberate and engage with each other before reaching a decision. In this way, judges on diverse benches are “exposed to new perspectives and forced to reevaluate their initial conclusions,”³⁵⁷ conclusions that may be improperly influenced by their preexisting ideas and assumptions about the validity of that evidence.

Interestingly, the jurists themselves seem to hold strong—yet conflicting—views on the matter. United States Supreme Court Justice Sandra Day O’Connor once expressed her view that “there’s simply no empirical evidence that gender differences lead to discernable differences in rendering judgment,”³⁵⁸ while Supreme Court of Canada Justice Bertha Wilson felt that “gender difference has been a significant factor in judicial decision-making, particularly in the areas of tort law, criminal law, and family law.”³⁵⁹ Many scholars agree with Justice Wilson, including the hundreds of legal academics involved in the United States Feminist Judgments Project, a “global collaboration of hundreds of feminist law professors who reimagine and rewrite key judicial decisions from a feminist perspective.” This project aims to show that the law is not neutral but is instead “deeply influenced by the perspectives of those who are appointed to interpret it.”³⁶⁰ Research on the impact of diversity on decision-making strongly supports this view and suggests that a diverse judiciary will be better equipped to question and more carefully analyze the precedential value of cases admitting now-discredited forensic evidence.

B. More Judicial Education in Science & Decision-making

The 2009 NRC report recommended that judicial education organizations establish continuing legal education programs for judges and many new training programs have been developed by organizations including the American Bar Association, the Federal Judicial Center, the National Courts and Sciences institute, and the National Judicial College.³⁶¹ For example, the Federal Judicial Center offers science education programs as part of its continuing judicial education program³⁶² and the National Courts and Sciences Institute (NCSI) has a mission “to enhance the

³⁵⁶ See Charlan Jeanne Nemeth, *Differential Contributions of Majority and Minority Influence*, 93 PSYCHOL. REV. 23, 28 (1986) (noting that individuals “exposed to persistent majority views tend toward convergence of thinking and to an unreflective acceptance of the majority position” but that “individuals exposed to persistent minority views are actually better decision makers in that they attend to more aspects of the situation, and they examine and reexamine premises.”).

³⁵⁷ Jason Iuliano & Avery Stewart, *The New Diversity Crisis in the Federal Judiciary*, 84 TENN. L.R. 247, 258 (2016) (noting that “[d]eep-level diversity increases the likelihood that judges will confront diverse values, attitudes, and experiences during their discussions.”).

³⁵⁸ SANDRA DAY O’CONNOR, *THE MAJESTY OF THE LAW: REFLECTIONS OF A SUPREME COURT JUSTICE* 197 (Random House Publishing 2004).

³⁵⁹ Bertha Wilson, *Will Women Judges Really Make a Difference?* 28 OSGOOD HALL LAW JOURNAL 507, 512 (1990).

³⁶⁰ US Feminist Judgments Project available at <https://law.unlv.edu/us-feminist-judgments> (last visited Dec. 10, 2024).

³⁶¹ Brandon L. Garrett et al., *Judges and Forensic Science Education: A National Survey*, 321 FORENSIC SCI. INT’L 1, 2 (2021).

³⁶² Federal Judicial Center, Judicial Programs, <https://www.fjc.gov/education/education-programs> (last visited Aug. 10, 2024).

capacity of courts to resolve complex cases involving novel scientific and technical evidence.”³⁶³ Many individual states offer similar training.³⁶⁴

Moreover, the vast majority of judges agree with the need for additional training in forensic science. A recent study of 164 judges from 39 different states who attended past trainings at the National Judicial College found that 92% of respondents reported that they should receive continuing education on forensic science evidence.³⁶⁵ And the NCSI’s decadal survey of judicial expectations regarding novel scientific evidence found that many state and territorial court judges predict that they will see cases involving comparative forensics in 2020-2030³⁶⁶ are interested in receiving additional education about forensic science.³⁶⁷

Various studies have shown the impact of judicial training on science evidence, including a greater familiarity with the statistical methods that underlie forensic science evidence, and a stronger belief that it is the judge’s responsibility to prevent “junk science” from being admitted at trial.³⁶⁸ Judges with more extensive training were also more likely to be aware of resources that could help them critically evaluate forensic evidence. As the authors noted, “judicial training may not only help judges, but help them to help themselves.”³⁶⁹

In addition to training on the science itself, however, judges would benefit from training on recognizing and addressing the impact of heuristics and biases on their decision-making.³⁷⁰ First, awareness of one’s own vulnerability to bias and to the bias blind spot is important when seeking

³⁶³ National Courts and Sciences Institute, <https://www.courtsandsciences.org/home> (last visited Aug. 10, 2024).

³⁶⁴ See e.g., Barbara Parker Hervey, A Texas Approach to the Intersection of Science and the Law (Nov. 1, 2017), available at https://www.americanbar.org/groups/judicial/publications/judges_journal/2017/fall/texapproach-intersectscience-and-law/; Milt Nuzum, *Science Education for Judges in Ohio*, 56 JUDGES J. 23 (2017).

³⁶⁵ Brandon L. Garrett et al., *Judges and Forensic Science Education: A National Survey*, 321 FORENSIC SCI. INT’L 1, 10 (2021) (“Judges reported high levels of interest in training on topics such as digital, DNA, toxicology, firearms, and fingerprint evidence. Judges showed lower levels of training interest in shoeprint and bitemark evidence.”).

³⁶⁶ NATIONAL COURTS & SCIENCES INSTITUTE, JUDGES’ FORECASTS & PREFERENCES FOR MANAGING SCIENTIFIC EVIDENCE IN COMPLEX CASES 2020-2030: A REPORT OF A SURVEY OF STATE & TERRITORIAL COURTS 37 (Oct. 17, 2020),

https://static1.squarespace.com/static/589de119197aea0ba4c0aba1/t/5f96eac3598862662f02aeb4/1603726028013/NCSI_Survey_2020-2030.pdf [hereinafter NCSI SURVEY]. On a scale of 0 (low probability) to 10 (high probability), the average rating score was 5.59, indicating a moderately high likelihood.

³⁶⁷ *Id.* at 34. A total of 738 judges made 2280 selections. The most frequently expressed preference, with 412 nominations, was principles of forensic science.

³⁶⁸ Brandon L. Garrett et al., *Judges and Forensic Science Education: A National Survey*, 321 FORENSIC SCI. INT’L 1, 4 (2021). (“However, judicial training was associated with 16 estimated rates of ruling forensic science evidence inadmissible. Specifically, judges who endorsed more extensive training in forensic science evidence indicated that they ruled such evidence to be inadmissible at greater rates.”).

³⁶⁹ Brandon L. Garrett et al., *Judges and Forensic Science Education: A National Survey*, 321 FORENSIC SCI. INT’L 1, 6 (2021).

³⁷⁰ For an excellent overview of the implications of cognitive bias for legal decision-making, see Gary Edmond & Kristy A. Martire, *Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making*, 82 MODERN L.R. 603 (2019); see also BRIAN BARRY, HOW JUDGES JUDGE: EMPIRICAL INSIGHTS INTO JUDICIAL DECISION-MAKING 184 (Routledge 2021). Such training does not appear to be widespread, however. For example, the National Judicial College offers just two sessions of “Decision-making” in its 2024-2025 schedule of courses, available at <https://www.judges.org/courses/> (last visited Dec. 8, 2024).

to reduce bias and improve decision-making regarding the admissibility of forensic evidence.³⁷¹ Next, while awareness of bias is generally insufficient to remove bias from decision-making, it can be the first step toward debiasing. Indeed, cognitive debiasing is not a single event, but instead proceeds through a variety of stages “from a state of lack of awareness of bias, to awareness, to the ability to detect bias, to considering a change, to deciding to change, then initiating strategies to accomplish change, and finally, maintaining the change.”³⁷² And while cognitive debiasing is far from an exact science, it is a skill that can be learned, and one that studies from other domains—particularly the behavioral sciences and medicine—suggest would assist judges in better evaluating long-standing but scientifically flawed forensic techniques.³⁷³

Debiasing techniques—active steps known as cognitive forcing strategies—are designed to mitigate the effects of cognitive biases and may hold potential for reducing bias in judges evaluating the admissibility of forensic evidence.³⁷⁴ Such interventions are quite common outside of the legal domain. Medical researchers, for example, have developed procedures like double-blind clinical trials to insulate them from the various biases that may affect decision-making.³⁷⁵ And while early research on debiasing failed to show much promise in minimizing cognitive biases,³⁷⁶ more recent studies are more optimistic about the impact of debiasing-training.³⁷⁷ Although there are many different ways of categorizing debiasing techniques, Baruch Fischhoff’s framework of person-related and task-related debiasing strategies offers a structured approach to identify strategies that may help reduce cognitive bias in judicial decision-making, particularly in evaluating the admissibility of forensic evidence.³⁷⁸

Person-related debiasing involves interventions that improve the individual’s ability to recognize and mitigate biases. This type of debiasing draws from research on the benefits of education, thinking strategies, and more formal decision aids people can be taught to use.³⁷⁹ This approach

³⁷¹ Irene Scopelliti, *Bias Blind Spot: Structure, Measurement, and Consequences*, 61 *MANAGEMENT SCI.* 2468, 2482 (2015).

³⁷² Pat Croskerry et al., *Cognitive Debiasing 2: Impediments To and Strategies for Change*, 22 *BMJ QUALITY & SAFETY* ii65 (2013).

³⁷³ Pat Croskerry et al., *Cognitive Debiasing 2: Impediments To and Strategies for Change*, 22 *BMJ QUALITY & SAFETY* ii65 at Table 1 and Table 2 (2013). (summarizing strategies for cognitive debiasing in clinical settings).

³⁷⁴ Pat Croskerry et al., *Cognitive Debiasing 2: Impediments To and Strategies for Change*, 22 *BMJ QUALITY & SAFETY* ii65 at Table 1 and Table 2 (2013). (summarizing strategies for cognitive debiasing in clinical settings); Richard P. Larrick, *Debiasing in* *BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING* (Derek J. Koehler & Nigel Harvey eds.) (2004).

³⁷⁵ See e.g. Simon J. Day & Douglas G. Altman, *Blinding in Clinical Trials and Other Studies*, 321 *BMJ* 504, 504 (2000).

³⁷⁶ See e.g. Baruch Fischhoff, *Debiasing in* *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 440 (Daniel Kahneman, et al. eds.) (1982) (reviewing studies of methods to reduce the hindsight bias and the overconfidence bias and finding that “[b]oth biases have proven moderately robust, resisting attempts to interpret them as artifacts and eliminate them by “mechanical” manipulations, such as making subjects work harder.”).

³⁷⁷ See e.g. Anne-Laure Sellier, *Debiasing Training Improves Decision Making in the Field*, 30 *PSYCHOL. SCI.* 1371, 1371 (2019) (describing “promising evidence that debiasing-training effects transfer to field settings and can improve decision making in professional and private life.”).

³⁷⁸ Baruch Fischhoff, *Debiasing in* *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 424 (Daniel Kahneman, et al. eds.) (1982) (observing that “Debiasing procedures may be most clearly categorized according to their implicit allegation of culpability. The most important distinction is whether responsibility for biases is laid at the doorstep of the judge, the task, or some mismatch between the two.”).

³⁷⁹ Jack B. Soll et al., *A Users Guide to Debiasing in* *BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING* 925 (Gideon Keren & George Wu eds.) (2015).

assumes that the situation the decisionmaker finds herself in is relatively fixed and that if given the proper training, knowledge and tools, the decisionmaker can be helped to overcome her biases. For instance, cognitive training programs can educate individuals about common biases, such as confirmation bias or anchoring, and teach them strategies to counteract these tendencies. In this “modify the decisionmaker approach,” strategies often emphasize critical thinking and probabilistic reasoning, helping decision-makers assess evidence more accurately.³⁸⁰

Person-related debiasing techniques could be used to improve judges’ cognitive skills and awareness of biases, thereby improving their evaluations of the reliability of forensic evidence. Educational interventions, for example, could be used to train judges to recognize specific biases that influence their evaluations of forensic evidence. One study found that economics professors, as opposed to those in the humanities or biology, were more likely to ignore sunk costs in their personal decisions.³⁸¹ Judges untrained in the impact of cognitive bias on decision-making may be unaware of how biases such as anchoring and confirmation bias influence their evaluation of scientific evidence. For example, if a judge is presented with a forensic expert’s high-confidence testimony, they might anchor their judgment on this assertion, even if the underlying methodology is flawed. Educating judges about common biases can help them to more critically evaluate such evidence. This training could also include instruction on statistical reasoning, enabling judges to critically evaluate probabilistic claims, error rates, and the significance of validation studies.³⁸² For example, judges reviewing ballistics matching evidence could learn to better distinguish between accurate probabilistic conclusions and misleading expert testimony that claims a match “to an exact statistical certainty,”³⁸³ or “to the exclusion of every other firearm in the world.”³⁸⁴

Specific debiasing forcing strategies like “consider the opposite”—also known as “analysis of competing hypotheses”—may also reduce cognitive bias in judges. This technique asks decisionmakers to “think of alternative hypotheses early in the hypothesis-evaluation process.”³⁸⁵ A recent study employing this method in police investigations in a missing person’s case found that investigators who were instructed to actively consider the opposite (of the guilt of the suspect, in this case) reported more hypotheses in general, as well as more non-criminal hypotheses about what may have happened to the missing person.³⁸⁶ As the authors noted, the debiasing strategy “can protect investigators from unconsciously favoring evidence that supports their initial suspicion and incites the use of confirming investigation strategies.”³⁸⁷ Other person-related debiasing

³⁸⁰ Jack B. Soll et al., *A Users Guide to Debiasing in* BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING 926 (Gideon Keren & George Wu eds.) (2015).

³⁸¹ Richard P. Larrick et al., *Teaching the Use of Cost-Benefit Reasoning in Everyday Life*, 1 PSYCHOL. SCI. 362, 365 (1990).

³⁸² See e.g. Geoffrey T. Fong et al., *The Effects of Statistical Training on Thinking about Everyday Problems*, 18 COGNITIVE PSYCHOL. 253, 262 (1986) (finding that “statistical training serves to enhance the use of statistical principles in reasoning.)

³⁸³ See e.g. *United States v Montiero*, 407 F. Supp 351 (D. Mass 2006).

³⁸⁴ See e.g. *United States v. Green*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005).

³⁸⁵ Richard S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN’L. PSYCHOL. 175, 211 (1998).

³⁸⁶ Ivar Fahsing et al., *Have You Considered the Opposite? A Debiasing Strategy for Judgment in Criminal Investigation*, 96 POLICE J. 45, 49 (2023).

³⁸⁷ Ivar Fahsing et al., *Have You Considered the Opposite? A Debiasing Strategy for Judgment in Criminal Investigation*, 96 POLICE J. 45, 54 (2023). Other studies, however, have reached different results. See e.g. Enid

techniques include slowing down strategies, where decisionmakers learn to take planned-time outs to “stop and regroup...remove distractions or focus more intently,” allowing for reduced cognitive load and better decision-making,³⁸⁸ and simulation training, where decisionmakers experience a simulation involving a difficult decision that includes a “cognitive error trap,” allowing them to learn about identifying and remediating common cognitive errors.³⁸⁹

In the courtroom, a judge using a “consider the opposite” technique in a case involving microscopic hair evidence, for example, might reflect on cases where that type of evidence led to wrongful convictions, or ask whether contextual information like a suspect’s criminal history could have influenced the forensic expert testifying in the case. Similarly, by considering alternative explanations or hypothetical testimony from a defense expert, judges could critically assess the credibility and potential biases of the presented testimony. This reflective process would encourage judges to question the presenting expert’s methodology and question the reliability of the underlying scientific technique instead of simply relying on past precedent to admit the evidence. Similarly, a judge undergoing simulation training might be asked to proceed over a *Daubert* hearing on the admissibility of blood spatter evidence. When presented with a confident expert witness with a long career in policing and a short time period in which to make a decision, the judge may initially admit the evidence without appropriately applying the full range of *Daubert* factors. An instructor would then reveal the cognitive errors—authority bias and confirmation bias in this case—that might have influenced their decision. In a second simulation and armed with this insight about the effect of bias on decision-making, the judge would be better prepared to more rigorously apply the *Daubert* factors and question the underlying methodology of the technique before reaching a conclusion.

In contrast, task-related debiasing modifies the decision-making environment or the way information is presented, reducing the likelihood of bias without requiring significant cognitive effort from the decision-maker.³⁹⁰ This type of “decision architecture” attempts to “alter the environment to provide a better match for the thinking that people naturally do when unaided or, alternatively, to encourage better thinking.”³⁹¹ For example, research suggests that when people are hungry, they have reduced mental and emotional resources, which may cause them to rely more heavily on cognitive shortcuts to reach faster decisions. A well-known study of Israeli parole-board decisions found that the percentage of favorable rulings decreased from approximately 65% to nearly 0% from the first ruling to the last ruling within each session but rebounded to a 65% favorable rate following a food break. The study’s authors attempted this discrepancy to “mental depletion,” and concluded that “when judges make repeated rulings, they show an increased

Maegherman et al., *Test of the Analysis of Competing Hypotheses in Legal Decision-Making* 35 APPL. COGNITIVE PSYCHOL. 52 (2021) (finding “no difference between those who received the ACH instructions and those who did not receive the ACH instructions in terms of perception of guilt, nor in their search for further information.”).

³⁸⁸ Carol-Anne Moulton et al., *Slowing Down to Stay Out of Trouble in the Operating Room: Remaining Attentive in Automaticity*, 85 ACADEMIC MED. 1571, 1576 (2010); Carol-Anne E. Moulton et al., *Slowing Down When You Should: A New Model of Expert Judgment*, 82 ACADEMIC MED. S109 (2007).

³⁸⁹ William F. Bond et al., *Using Simulation to Instruct Emergency Medicine Residents in Cognitive Forcing Strategies*, 79 ACADEMIC MED. 438 (2004).

³⁹⁰ Richard P. Larrick, *Debiasing in* BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING (Derek J. Koehler & Nigel Harvey eds.) (2004).

³⁹¹ Jack B. Soll et al., *A Users Guide to Debiasing in* BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING 926 (Gideon Keren & George Wu eds.) (2015).

tendency to rule in favor of the status quo.”³⁹² We may be able to improve the decision-making environment by simply encouraging judges to avoid making important decisions when they are tired, hungry or angry.

Task-related debiasing techniques could also help improve judicial decision-making by introducing decision aids to assist judges when they evaluate complex scientific evidence. For example, structured decision aids, such as checklists, have been shown to improve judgment by guiding individuals through systematic evaluations of evidence.³⁹³ Technological tools like decision support systems also fall under this category, and have been widely adopted in the healthcare context to enhance clinical decisions and patient outcomes.³⁹⁴ One study of decisions made by judges in trademark cases found that although the test to assess consumer confusion in trademark cases contains over eight factors, most judges rely on just two—similarity and proximity of the commodities in question. The author concluded that judges employed heuristics “to short-circuit the multifactor test.”³⁹⁵ Another study similarly found that judges in England and Wales experience “factor overload,” leading them to rely on intuitive decision-making and “default to a coin flip.”³⁹⁶

In the case of forensic evidence, judges could use admissibility checklists that include prompts to evaluate whether the method has been peer-reviewed, its known error rate, and whether it has been independently validated. Similarly, when judges are faced with the admissibility of a particular kind of forensic evidence, they could consult a decision support system that provides an up-to-date summary of known error rates, validation studies, and expert criticisms. This practice would help ensure that judges base their rulings on current scientific consensus rather than outdated practices or unreliable precedent. Both of these techniques could help ensure that judges consider all relevant factors and reduce the risk of heuristic shortcuts, such as relying solely on a technique’s general acceptance in the field and largely ignoring other *Daubert* factors, or by simply relying on past precedent to justify the admission of unreliable forensic evidence.

Panel reviews of controversial forensic methods are another example of a task-related debiasing strategy. For example, a panel of judges with specialized training in scientific evidence could collectively evaluate the admissibility of forensic evidence whose reliability has been questioned. Because group decision-making processes help reduce individual biases by incorporating diverse

³⁹² Shai Danziger et al., *Extraneous Factors in Judicial Decisions*, 108 PNAS 6889, 6892 (2011); but see Andreas Glöckner, *The Irrational Hungry Judge Effect Revisited: Simulations Reveal that the Magnitude of the Effect is Overestimated*, 11 JUDGMENT & DECISION MAKING 601, 607 (2016) (suggesting that the dramatic drop in the Israeli study “does not conclusively indicate bias or error in judicial decision making” as parole board decisions are unlikely to be randomly distributed across a session, and because favorable rulings take longer than unfavorable rulings, judges may schedule favorable cases earlier in sessions.).

³⁹³ See John W. Ely et al., *Checklists to Reduce Diagnostic Errors*, 86 ACADEMIC MEDICINE 307, 308 (2011) (noting that “[c]hecklists could help us resist the biases and failed heuristics that lead to diagnostic errors.”).

³⁹⁴ See e.g. Reed T. Sutton et al., *An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies For Success*, 3 DIGITAL MEDICINE 1 (2020); Malek Elhaddad & Sara Hamam *AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential*, 16 CUREUS 1 (2024).

³⁹⁵ Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CAL. L.R. 1581, 1614 (2006).

³⁹⁶ Inbar Levy, *Simplifying Legal Decisions: Factor Overload in Civil Procedure Rules*, 41 MELBOURNE L.R. 727, 751 (2017) (concluding that “a long list of factors might have detrimental effects in legal decision-making.”).

perspectives and promoting critical deliberation,³⁹⁷ this approach could be particularly valuable in cases where judges face conflicting expert testimony or unclear scientific consensus. Similarly, replacement is a forcing strategy where a decisionmaker who may be biased in favor of a particular outcome is replaced with a disinterested decision-maker or groups of decisionmakers.³⁹⁸ In the courtroom, a version of this debiasing strategy could be accomplished by employing more independent scientific advisors under Rule 702 to assist judges with scientific issues, a reform that has been suggested by a number of legal scholars.³⁹⁹ These independent experts can help judges navigate complex scientific evidence, providing objective evaluations that would allow judges to make more informed decisions that align with the requirements of Rule 702. Other task-related debiasing strategies include software tools that analyze the statistical validity of forensic methods to assist judges in identifying unreliable science before it reaches the jury, and predictive modeling based on data from prior cases, which could flag forensic methods that have been frequently overturned or discredited, helping judges make more informed admissibility decisions.⁴⁰⁰

Combining person- and task-related techniques would offer the most robust solutions for mitigating biases. In a case involving the admissibility of voiceprint analysis, for example, judges could use their training in the statistical underpinnings of forensic methods to recognize the potential limitations of this method of forensic matching analysis and carefully scrutinize its error rates and validation. And during a *Daubert* hearing, the judge could use a decision aid like an admissibility checklist to ensure a systematic evaluation of the evidence, asking whether the method has undergone peer review, what the error rates are, and whether it has been accepted in the scientific community. This combination of knowledge from training (person-related) and structured evaluation (task-related) could help the judge critically assess the method's reliability, reducing the risk of admitting unreliable evidence.

While promising, these techniques of course face limitations. Educational programs may require significant time and resources, and task-related tools like checklists might be perceived as burdensome or infringing on judicial discretion. Moreover, because many biases are applied unconsciously, if judges are unaware of them, they may fail to recognize the need for debiasing in the first place or be resistant to reforms seeking to address unconscious bias. And even when people become aware of the impact of cognitive bias on their decision-making, an awareness of bias does not always translate into successful correction, particularly under stress or time pressure. Tellingly, even Daniel Kahneman, widely considered the father of behavioral science, admitted

³⁹⁷ See *infra* at Part ____.

³⁹⁸ Gary Edmond & Kristy A. Martire, *Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making*, 82 MODERN L.R. 603, 658 (2019).

³⁹⁹ See e.g., Jonathan J. Koehler, *How Trial Judges Should Think about Forensic Science Evidence*, 102 JUDICATURE 28, 35 (Spring 2018) (noting that “judges might consider enlisting disinterested scientists who have relevant methodological expertise.”); Jean Sternlight, *Justice in a Brave New World?*, 52 CONN. L.R. 213, 257 (2020) (suggesting “a panel of neutral highly credentialed scientific experts, paid for by tax dollars, charged with examining certain kinds of scientific evidence pertaining to claims brought in a particular jurisdiction.”); David DeMatteo et al., *Expert Evidence, The (Unfulfilled) Promise of Daubert*, 20 PSYCHOL. SCI. PUBLIC INTEREST 129, 132 (2019); James R. Steiner-Dillon, *Epistemic Exceptionalism*, 52 INDIANA L.R. 207, 252 (2019); but see NCSI Survey at 14 (finding that many respondents did not feel a scientific staff officer would be helpful to the court, with 51.3% rating the helpfulness to judges low to moderate. Many judges expressed concerns about ex parte communications and the appearance of bias.).

⁴⁰⁰ See, e.g. Sandra Eloranta & Magnus Boman, *Predictive Models for Clinical Decision Making: Deep Dives in Practical Machine Learning*, 292 J. INTERN. MED. 278 (2022).

that he was prone to cognitive biases, noting that “I’ve [studied biases and errors of judgement] for 25 years now, and I think I haven’t improved at all.”⁴⁰¹

Notwithstanding these challenges, however, we know that judges—like all decisionmakers—are susceptible to cognitive bias. And while we may not be able to claim with certainty that a particular decision is a direct result of a particular cognitive bias, we can confidently say that these biases at least contribute to the ongoing failure of courts to properly apply *Daubert* to the admission of forensic evidence. It is therefore not the intent of this Article to criticize the hard work of judges making difficult decisions on admissibility in an environment of rapid scientific change. Instead, this Article means to bring awareness of the unconscious biases that affect all decisionmakers, including judges, because when people are aware of their own biases, they are more likely to engage in bias reduction strategies and to respond to training on debiasing.⁴⁰² After all, as Pat Croskerry and colleagues note, “[o]ngoing cognitive debiasing is arguably the most important feature of the critical thinker and the well-calibrated mind.”⁴⁰³

CONCLUSION

There is not a simple solution to the problem of flawed forensic science, and it is not just courts that bear the responsibility for keeping unreliable forensic evidence out of the courtroom. The prosecutors and defence attorneys litigating these cases also have a professional obligation to gain the scientific competence necessary to recognize the weaknesses in many types of forensic methods and to avoid proffering invalid opinions of unqualified “experts.” The forensic science community has a similar responsibility to address deficiencies in forensic methods, develop objective measures of assessment and interpretation when those measures can be established, and discard other techniques that will never meet such standards. All of these steps are crucial and urgent.

Ultimately, however, the judge is the gatekeeper and the final check on the admissibility of forensic evidence in their own courtrooms. So while acknowledging bias is an important goal for all decisionmakers, it is especially important when it comes to reducing the use of faulty forensic evidence. Judges must be willing to consider the current state of the science and to revisit and correct the old cases admitting forensic evidence simply because it has long been assumed reliable. And an awareness of the cognitive biases that might prevent judges from explicitly and transparently engaging with the reliability of that evidence may assist them in this important task.

⁴⁰¹ Cass Sunstein, *Rediscovering the Unconscious*, HARVARD MAGAZINE (2014), available at <https://www.harvardmagazine.com/2014/02/rediscovering-the-unconscious#:~:text=In%20his%20own%20life%2C%20though,%20improved%20at%20all.%E2%80%9D>

⁴⁰² Cynthia McPherson Frantz, *I AM Being Fair: The Bias Blind Spot as a Stumbling Block to Seeing Both Sides*, 28 BASIC & APPLIED SOC. PSYCHOL. 157, 158 (2006). (“When we have accurate theories about what influences our judgment, we can effectively correct a bias if we were motivated to do so. If we remain unaware of a bias, we cannot correct for it.”) Many organizations, including the American Bar Association, offer continuing education training on cognitive bias and the ways in which unconscious bias can affect decision-making. Mary Smith et al., *Addressing Bias Among Judges* (Sept. 14, 2023), <https://statecourtreport.org/our-work/analysis-opinion/addressing-bias-among-judges>.

⁴⁰³ Pat Croskerry et al., *Cognitive Debiasing 2: Impediments To and Strategies for Change*, 22 BMJ QUALITY & SAFETY ii65, ii65 (2013) (examining the “major strategies that have been developed in the social sciences and in medicine to achieve cognitive and affective debiasing.”)