

# NOTES

## OVERCOMING *DAUBERT*'S SHORTCOMINGS IN CRIMINAL TRIALS: MAKING THE ERROR RATE THE PRIMARY FACTOR IN *DAUBERT*'S VALIDITY INQUIRY

MUNIA JABBAR\*

*Daubert v. Merrell Dow Pharmaceuticals, Inc. and its progeny provide the federal standard for the admissibility of all expert evidence, including forensic evidence, that is proffered in criminal trials. The standard measures the validity of expert evidence through a flexible four-factor inquiry. Unfortunately, in the criminal context, Daubert fails to promote the goals of trial outcome accuracy and consistency, resulting in tragically unfair outcomes for criminal defendants. This Note proposes a doctrinal tweak that shifts the costs of admitting forensic evidence to the prosecution and promotes criminal justice goals. First, there should be a high presumption against the admission of forensic evidence that must be rebutted with a clear and convincing showing of its validity. Second, the Daubert validity inquiry needs to be reformulated so that the forensic methodology's "error rate" factor is the primary (and if possible, only) factor the court considers. Third, the error rate should be defined as the lab-specific error rate. The Note ends by considering further possible ways to specify the definition of "error rate" to better promote criminal justice goals.*

### INTRODUCTION

Jeffrey Pierce was exonerated after spending fifteen years in prison for a rape he did not commit.<sup>1</sup> Despite a plausible alibi, Pierce was convicted largely due to the hair analysis conducted by Oklahoma City police chemist Joyce Gilchrist.<sup>2</sup> A preliminary Federal Bureau of Investigation study of eight cases involving Gilchrist found that, in five of them, she had overstepped "the acceptable limits of forensic sci-

---

\* Copyright © 2010 by Munia Jabbar. J.D., 2010, New York University School of Law; B.A., 2006, Harvard University. I am immensely grateful to many people for their input and help. I would like to thank Professor Daniel Rubinfeld for providing me with crucial feedback on an early draft. I am also thankful for feedback and advice from Kristen Richer and Shawn Ledingham in helping me develop this piece. Finally, Seth Harp, Whitney Cork, Kirstin Kerr O'Connor, and Sallie Kim have been very patient and generous in editing this Note.

<sup>1</sup> Belinda Luscombe, *When the Evidence Lies*, TIME, May 13, 2001, at 38.

<sup>2</sup> *Id.* at 39.

ence” by exaggerating the certainty and accuracy of her results.<sup>3</sup> The Oklahoma governor ordered a review of thousands of cases in which Gilchrist was involved.<sup>4</sup> By the time the review was implemented, however, eleven of the defendants Gilchrist helped convict had already been put to death.<sup>5</sup>

Gilchrist's careless methods and testimony had gone unaddressed for twenty-one years.<sup>6</sup> Her misbehavior is clearly disturbing,<sup>7</sup> but it is also alarming that judges continually allowed the crime lab's misleading results and testimony to be admitted at all. The rules for the admission of expert evidence, including forensic evidence, were such that inaccurate but persuasive evidence could be admitted in high-stakes criminal trials.

In the federal system and in about half of the states (including Oklahoma, where Pierce was convicted),<sup>8</sup> the admissibility standard for expert evidence is based on or influenced by the Supreme Court's ruling in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,<sup>9</sup> which interprets Federal Rule of Evidence 702.<sup>10</sup> Under *Daubert*, judges act as gatekeepers to evaluate whether the scientific or technical methodology underlying a proffer of expert evidence is “reliable” enough to be admitted.<sup>11</sup> The “central operating construct” of this evaluation is an inquiry into the scientific validity of the methodology.<sup>12</sup> This inquiry looks at four flexible factors: (1) testability of the underlying methodology, (2) peer review and publication of that methodology,

---

<sup>3</sup> *Id.* at 38 (“Gilchrist got convictions by matching hair samples with a certainty other forensic scientists found impossible to achieve.”). For Pierce's suit against Gilchrist and Oklahoma City, see *Pierce v. Gilchrist*, No. Civ-05-1519-C, 2007 U.S. Dist. LEXIS 3191 (W.D. Okla. Jan. 16, 2007).

<sup>4</sup> Luscombe, *supra* note 1, at 38.

<sup>5</sup> *Id.*

<sup>6</sup> *Id.* at 39.

<sup>7</sup> For instance, Gilchrist concealed and altered laboratory reports that could have exonerated defendants. Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1, 77 (2009).

<sup>8</sup> Alice B. Lustre, Annotation, *Post-Daubert Standards for Admissibility of Scientific and Other Expert Evidence in State Courts*, 90 A.L.R.5th 453, 480–519 (2001). The other states include Alaska, Arkansas, Colorado, Connecticut, Delaware, Idaho, Indiana, Iowa, Kentucky, Louisiana, Maine, Montana, Nebraska, New Mexico, North Carolina, Ohio, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Vermont, West Virginia, and Wyoming. *Id.*

<sup>9</sup> 509 U.S. 579 (1993).

<sup>10</sup> FED. R. EVID. 702. Though *Daubert* originally interpreted only the federal rule, the standard has also been adopted by many states. *E.g.*, *People v. Shreck*, 22 P.3d 68, 77–78 (Colo. 2001) (adopting flexible expert evidence standard for Colorado's Rule 702 equivalent and referring to *Daubert* factors as permissible, but not required, to consider).

<sup>11</sup> *Daubert*, 509 U.S. at 592.

<sup>12</sup> DAVID L. FAIGMAN ET AL., 1 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 15 (2d ed. 2002).

(3) error rate, and (4) general acceptance by the relevant scientific community.<sup>13</sup>

As seen in the *Pierce* case, relying on the *Daubert* standard does not always produce ideal results in criminal adjudication. There are three reasons for this trend. First, resource-constrained defense attorneys rarely manage to challenge forensic evidence proffered by the prosecution successfully,<sup>14</sup> a fact that opens the door to more misleading prosecutorial evidence and increases the likelihood of false convictions.<sup>15</sup> Second, there is a high risk that juries will place unfounded emphasis on forensic evidence, thus exacerbating the risk of unfair prejudice. Third, excessive flexibility in the *Daubert* validity factors has led to inconsistent application of the *Daubert* standard and unpredictable admission of expert evidence across jurisdictions.<sup>16</sup>

Many have offered administrative and legislative solutions to the problem of unreliable expert testimony,<sup>17</sup> while others have also proposed various doctrinal adjustments.<sup>18</sup> Some commentators have

<sup>13</sup> See *Daubert*, 509 U.S. at 593–95 (discussing factors).

<sup>14</sup> Robert E. Bernstein, *Expert Witnesses, Adversarial Bias, and the (Partial) Failure of the Daubert Revolution*, 93 IOWA L. REV. 451, 460–61 (2008).

<sup>15</sup> See Garrett & Neufeld, *supra* note 7, at 33–34 (discussing risks of prosecutors misleading jurors due to expert testimony resource imbalances).

<sup>16</sup> Cassandra H. Welch, *Flexible Standards, Deferential Review: Daubert's Legacy of Confusion*, 29 HARV. J.L. & PUB. POL'Y 1085, 1094 (2006).

<sup>17</sup> See, e.g., Elizabeth L. DeCoux, *The Admission of Unreliable Expert Testimony Offered by the Prosecution: What's Wrong with Daubert and How To Make It Right*, 2007 UTAH L. REV. 131, 163 (advocating for trial courts or special masters to review and address pertinent scientific articles); Paul C. Giannelli, *The Abuse of Scientific Evidence in Criminal Cases: The Need for Independent Crime Laboratories*, 4 VA. J. SOC. POL'Y & L. 439, 469–73 (1997) (recommending changes that would lead to independent crime labs); Peter J. Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggestions for Reform*, 95 AM. J. PUB. HEALTH S107, S111–13 (2005) (advocating for independent auditing and national accreditation system).

<sup>18</sup> See, e.g., Andrew Jurs, *Judicial Analysis of Complex & Cutting-Edge Science in the Daubert Era: Epidemiologic Risk Assessment as a Test Case for Reform Strategies*, 42 CONN. L. REV. 49, 84–88 (2009) (advocating for appointments of science consultants under modified Federal Rule of Evidence 706). Others have argued for doctrinal reforms that would increase the reliability of expert testimony, though not specifically in the *Daubert* context. E.g., Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence*, 95 CAL. L. REV. 721, 792–93, 795–97 (2007) (proposing several administrative and doctrinal solutions, including placing burden on government to show scientific technique's legitimacy and instituting threshold error rate for evidence admissibility). Much of the literature on the *Daubert* doctrine focuses on comparisons with the standard under *United States v. Frye*, 293 F. 1013 (D.C. Cir. 1923), for expert evidence admissibility rather than doctrinal reform efforts. E.g., Edward K. Cheng & Albert K. Yoon, *Does Frye or Daubert Matter? A Study of Scientific Admissibility Standards*, 91 VA. L. REV. 471, 503 (2005) (analyzing empirical differences in trial outcomes in states that have adopted *Frye* or *Daubert*); Paul C. Giannelli, *Admissibility of Scientific Evidence*, 28 OKLA. CITY U.L. REV. 1, 6–9 (2003) (discussing effects of *Daubert* and contrasting it with *Frye*). For a discussion of the *Frye* standard, see *infra* Part I.C.1.

argued for a presumption against admitting expert evidence by requiring such evidence to meet a threshold error rate.<sup>19</sup> I advance these arguments by proposing a novel doctrinal tweak of the *Daubert* standard for the admissibility of forensic evidence in criminal trials. In this Note, I assert that the error rate should be the primary factor in the validity inquiry under *Daubert*.

I start briefly with a proposition that others have argued: There should be a rebuttable presumption against admitting expert evidence that needs to be overcome by a clear and convincing showing of the evidence's validity. I then use statistical principles to propose a more rigid *Daubert* validity test that must be satisfied to rebut this presumption. Under my reformulated validity inquiry, judges first look at the specific error rate of the scientific or technical methodology as it is practiced in that instance; but I stop short of prescribing a threshold error rate for admissibility. I further propose that two of the other *Daubert* validity factors—general acceptance and peer review and publication—should come in as secondary validity factors if and only if relevant specific error rate information cannot be ascertained. I then analyze two possible ways to further define the error rate factor, both of which hinge on the type of error involved in the evidence proffer.<sup>20</sup>

This doctrinal adjustment would promote the broad goals of trial outcome accuracy and consistency that underlie all criminal trials. And generally, this would shift the costs of introducing forensic evidence toward the prosecution, closely aligning with fundamental principles of our criminal justice system: that the prosecution has the burden of proving guilt beyond a reasonable doubt, and that society prefers that guilty defendants go free rather than innocent defendants be unjustly condemned.

In Part I, I provide the foundation of my argument by introducing the goals of the criminal justice system, the applicable Federal Rules of Evidence, and the *Daubert* standard itself. In Part II, I discuss the problems with *Daubert* that have arisen in criminal cases. In Part III, I propose my solution for *Daubert* reform. I start by briefly justifying a presumption against the admissibility of expert evidence in light of *Daubert's* shortcomings. I then reformulate the *Daubert* validity

---

<sup>19</sup> *E.g.*, Murphy, *supra* note 18, at 796–97. I tentatively reject the idea of an across-the-board threshold error rate for admissibility. A detailed discussion of the merits of a hard threshold are beyond the scope of this Note.

<sup>20</sup> However, since most error rates are only applicable to forensic methodologies, my proposal is largely targeted at forensic evidence. For other types of expert evidence, my proposal leaves *Daubert* largely unchanged. My proposal is not limited to forensic evidence out of necessity—in theory, it works for all expert evidence proffers. However, limiting the discussion to forensic evidence lends clarity, and, because error rates are generally only available for forensic methodologies, my proposal is most effective for those proffers.

inquiry as focusing on the practice-specific error rate of the underlying methodology and discuss ways to clearly define the error rate so as to better serve criminal justice goals.

## I

### BACKGROUND: CRIMINAL JUSTICE GOALS, THE FEDERAL RULES OF EVIDENCE, AND THE *DAUBERT* STANDARD

Any system for expert evidence admissibility should endorse the two normative criminal justice goals of trial accuracy and consistency. Unfortunately, these goals are not well promoted by the current law consisting of Federal Rules of Evidence 702 and 403 and the *Daubert* doctrine that interprets them.

#### A. *Two Key Normative Goals of the Criminal Justice System*

There are two key goals of the criminal justice system that would be promoted by a reformulation of the *Daubert* standard.

The first goal is achieving an optimally low level of inaccurate trial outcomes.<sup>21</sup> Overall, one wants to avoid both false convictions and false acquittals,<sup>22</sup> but society places a higher priority on preventing false convictions.<sup>23</sup> This prodefendant stance is consistent with most normative conceptions of the criminal justice system.<sup>24</sup> In reformulating the *Daubert* standard, I will prioritize mitigating the risk of false convictions even at the cost of increasing the risk of false acquittals.<sup>25</sup>

The second criminal justice goal is predictability and consistency across cases.<sup>26</sup> Predictability is necessary for fairness in the criminal

---

<sup>21</sup> E.g., Vikramaditya S. Khanna, *Double Jeopardy's Asymmetrical Appeal Rights: What Purpose Do They Serve?*, 82 B.U. L. REV. 341, 357 n.75 (2002) (discussing relative social costs of favoring false acquittals to false convictions).

<sup>22</sup> *Id.*

<sup>23</sup> *Id.*

<sup>24</sup> For instance, the fact that the burden of proof necessary for a criminal conviction is set at “beyond a reasonable doubt” embodies the criminal justice system’s fundamental preference for avoiding false convictions. LARRY LAUDAN, *TRUTH, ERROR, AND CRIMINAL LAW: AN ESSAY IN LEGAL EPISTEMOLOGY* 29–30 (2006).

<sup>25</sup> This is in some sense a tradeoff. In setting a burden of proof of “beyond a reasonable doubt,” society is more concerned with preventing false convictions than false acquittals. *Id.* This tradeoff translates similarly to rules governing the admissibility of forensic evidence. More concern over false convictions than false acquittals should increase the barrier for the admission of expert evidence if, as is usually the case in criminal trials, it is proffered by the prosecution. Cf. Murphy, *supra* note 18, at 793 (“[T]he government should carry a burden of placing before the court continued evidence of a technique’s legitimacy.”).

<sup>26</sup> See Robert F. Schopp, *Verdicts of Conscience: Nullification and Necessity as Jury Responses to Crimes of Conscience*, 69 S. CAL. L. REV. 2039, 2100 (1996) (“[A]ny plausible institutional structure for the conventional public morality will include a system of law that

justice system as well as for establishing the rule of law consistent with the “fundamental principles of a liberal society.”<sup>27</sup> The *Daubert* standard should be made clearer in order to promote easy and consistent application. Consistency in the value of admitted evidence should tend to increase the consistency of trial outcomes.<sup>28</sup>

### B. Federal Rules of Evidence 702 and 403

The two Federal Rules of Evidence most relevant to the *Daubert* standard are Rule 702, which governs the admissibility of expert evidence, and Rule 403, which governs the admissibility of nearly all evidence by directing judges to weigh its probative value against any unfair prejudice it may create. *Daubert* directly interprets Rule 702 and implicitly addresses the interaction between Rules 403 and 702.<sup>29</sup>

#### 1. Rule 702: Expert Evidence Admissibility

Rule 702 governs the admissibility of expert evidence in the federal system.<sup>30</sup> The rule allows “scientific, technical, or other specialized knowledge” to be admitted if it is relevant.<sup>31</sup> Such evidence may be in the form of expert testimony.<sup>32</sup> The expert must be qualified, and her testimony may be “in the form of an opinion or otherwise.”<sup>33</sup>

*Daubert* interpreted Rule 702 to require an investigation into the validity of the testimony.<sup>34</sup> This requirement was incorporated into the 2000 amendment to Rule 702 so that “all types of expert testimony

---

can represent that conventional morality in a relatively consistent, predictable, and enforceable manner . . .”).

<sup>27</sup> *Id.*

<sup>28</sup> This is illustrated by the fact that the quality of expert evidence heard by the jury affects the jury’s final decision. See Garrett & Neufeld, *supra* note 7, at 33–34 (explaining that invalid expert testimony can “lead[ ] the jury to draw incorrect conclusions”).

<sup>29</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 595 (1993).

<sup>30</sup> Rule 702 states:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

FED. R. EVID. 702.

<sup>31</sup> *Id.*

<sup>32</sup> *Id.*

<sup>33</sup> *Id.*

<sup>34</sup> *Daubert*, 509 U.S. at 589–90; see also *infra* Part I.C.2 (outlining *Daubert* standard in light of Rule 702’s adoption).

present questions of admissibility for the trial court in deciding whether the evidence is reliable and helpful.”<sup>35</sup>

## 2. Rule 403: The Probative-Prejudicial Balance

Rule 403 can act as a constraint on Rule 702 by requiring the court to balance “the probative value of and need for the evidence against the harm likely to result from its admission.”<sup>36</sup> Rule 403 inherently subsumes Rule 702: After evidence is found sufficiently reliable under Rule 702, it can be excluded if it does not satisfy Rule 403.<sup>37</sup>

Rule 403 allows relevant evidence to be excluded “if its probative value is substantially outweighed by the danger of unfair prejudice.”<sup>38</sup> The Rule also weighs probative value against other considerations such as confusion of the issues and the risk of misleading the jury.<sup>39</sup> However, the Rule’s main function is to mitigate any unfair prejudice that may be caused by otherwise relevant evidence. There is no strict standard for when relevance is considered to be “substantially outweighed by . . . unfair prejudice”: Judges determine this on a case-by-case basis.<sup>40</sup> The key is that the prejudice must be of a significant enough degree to be considered unfair. Unfair prejudice can mean that the jury uses the evidence to assess the wrong factors or weighs it in an improper way.<sup>41</sup>

---

<sup>35</sup> FED. R. EVID. 702 advisory committee’s note to 2000 amendment.

<sup>36</sup> FED. R. EVID. 403 advisory committee’s note.

<sup>37</sup> FAIGMAN ET AL., *supra* note 12, at 26. The *Daubert* Court also stated that any expert evidence that is to be admitted under Rule 702 must also satisfy Rule 403. *See infra* notes 60–62 and accompanying text (explaining why Rule 403 must be considered in conjunction with Rule 702).

<sup>38</sup> FED. R. EVID. 403. Rule 403 states: “Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.” *Id.*

<sup>39</sup> *Id.* This rule carves out an exception to Rule 402, which says that any “relevant” evidence can be admissible. FED. R. EVID. 402. The term “relevant” is construed extremely broadly. *See* FED. R. EVID. 401 (defining “[r]elevant evidence” as evidence having “any tendency” to make fact more or less probable); *Daubert*, 509 U.S. at 587 (noting liberal relevance standard). Thus, Rule 401 is the starting point for admitting any piece of evidence. In general, relevant evidence will be admitted unless it falls into an exception such as Rule 403. *See* FED. R. EVID. 401 advisory committee’s note (explaining that some types of relevant evidence that fall under specific patterns, such as evidence that is unduly prejudicial, is subject to other exclusionary rules, such as Rule 403).

<sup>40</sup> Edward J. Imwinkelried, *Impoverishing the Trier of Fact: Excluding the Proponent’s Expert Testimony Due to the Opponent’s Inability To Afford Rebuttal Evidence*, 40 CONN. L. REV. 317, 351–52 (2007).

<sup>41</sup> FED. R. EVID. 403 advisory committee’s note (“‘Unfair prejudice’ within its context means an undue tendency to suggest decision on an improper basis, commonly, though not necessarily, an emotional one.”).

Unfortunately, Rule 403 does not always suffice to exclude overly prejudicial evidence.<sup>42</sup> Rule 403 “carries a strong presumption in favor of admissibility.”<sup>43</sup> This presumption is not necessarily appropriate for forensic evidence: Despite the high potential for prejudicial effect with such evidence, trial courts may not recognize this risk when making a Rule 403 determination. Defendants often do not have the resources to challenge prosecution proffers competently, and judges do not always have the information necessary to properly decide that certain forensic evidence should be excluded under Rule 403 for being too prejudicial.<sup>44</sup> Thus in practice, unfairly prejudicial forensic evidence can easily slip through the cracks of Rule 403.

### C. *The Daubert “Revolution”*

Before *Daubert*, the federal standard for the admission of expert evidence was established in *Frye v. United States*.<sup>45</sup> While the *Frye* inquiry required examining only one factor (general acceptance of underlying methodology), *Daubert* incorporates many additional factors to create a flexible test. However, under the *Daubert* standard, the system for admitting forensic evidence in criminal trials undermines criminal justice goals. The *Daubert* standard’s validity inquiry has been particularly contrary to such goals and is where I will focus my doctrinal reformulation.

#### 1. *The Frye Standard*

Until *Daubert*, the federal standard for the admissibility of expert evidence was articulated in the D.C. Circuit Court of Appeals’s brief opinion in *Frye*.<sup>46</sup> Written in 1923, the *Frye* opinion discussed whether to admit polygraph evidence.<sup>47</sup> The standard was a one-prong inquiry into the general acceptance of a scientific methodology: To be admissible, the scientific methodology underlying an expert evidence proffer “must be sufficiently established to have gained general

---

<sup>42</sup> Rule 403 does work sometimes to exclude unduly prejudicial evidence. Courts will explicitly or implicitly use Rule 403 to exclude polygraph evidence, for example. *E.g.*, *United States v. Sherlin*, 67 F.3d 1208, 1217 (6th Cir. 1995) (finding polygraph results excludable under Rule 403 because of “extremely dubious probative value”).

<sup>43</sup> *United States v. Hands*, 184 F.3d 1322, 1328 (11th Cir. 1999) (quoting *United States v. Church*, 955 F.2d 688, 703 (11th Cir. 1992)). This is evidenced by Rule 403’s requirement that the bases for exclusion must substantially outweigh the evidence’s probative value. FED. R. EVID. 403.

<sup>44</sup> See *infra* notes 91–99 and accompanying text (explaining how resource inequalities impede defense challenges to evidence admissibility).

<sup>45</sup> 293 F. 1013, 1014 (D.C. Cir. 1923).

<sup>46</sup> FAIGMAN ET AL., *supra* note 12, at 10.

<sup>47</sup> *Frye*, 293 F. at 1013.

acceptance in the particular field in which it belongs.”<sup>48</sup> This standard was heavily criticized for difficulties in delineating the scope of the scientific field, determining what comprises general acceptance by a field, and proving that such general acceptance exists.<sup>49</sup> The Supreme Court addressed these concerns when it decided *Daubert* in 1993.

## 2. *The Daubert Standard*

In *Daubert*, the Supreme Court held that the *Frye* standard had been superseded by the adoption of Rule 702.<sup>50</sup> The Court interpreted Rule 702 as providing a new standard for the admission of expert evidence.<sup>51</sup> The Court elaborated on the full scope of the standard in three cases: *Daubert*, *General Electric Co. v. Joiner*,<sup>52</sup> and *Kumho Tire Co. v. Carmichael*,<sup>53</sup> all of which arose in the context of civil litigation.

Under the *Daubert* standard, the trial judge serves a Rule 104(a) gatekeeper function<sup>54</sup> and must use a three-pronged test to determine admissibility: (1) the evidence must be relevant, (2) the expert presenting the evidence must be qualified, and (3) the evidence must be characterized by scientific validity.<sup>55</sup>

Procedurally, if a party opposes the admission of some evidence, it may file a *Daubert* motion. During a *Daubert* hearing, the judge makes “a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and whether that reasoning or methodology properly can be applied to the facts in

<sup>48</sup> *Id.* at 1014.

<sup>49</sup> *E.g.*, Paul Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 COLUM. L. REV. 1197, 1208–20 (1980); John McCabe, *DNA Fingerprinting: The Failings of Frye*, 16 N. ILL. U. L. REV. 455, 456–57 (1996). However, *Frye* is not a dead standard—sixteen states still adhere to it in some form. Lustre, *supra* note 8, at 453.

<sup>50</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 587–90 (1993).

<sup>51</sup> *Id.*

<sup>52</sup> 522 U.S. 136 (1997).

<sup>53</sup> 526 U.S. 137 (1999).

<sup>54</sup> *Daubert*, 509 U.S. at 592–93; *see also* FED. R. EVID. 104(a) (“Preliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court . . .”).

<sup>55</sup> *Daubert*, 509 U.S. at 587–93. The Court characterized this approach as a more liberal admission standard than *Frye*’s. *Id.* at 588. However, this was seemingly modified later in *General Electric Co. v. Joiner*, 522 U.S. 136, 142 (1997). Given the potential for expert evidence to be unfairly prejudicial, which had been acknowledged by Justice Blackmun himself in his *Barefoot v. Estelle* dissent, one questions Blackmun’s articulation of a permissive standard in *Daubert*. *Barefoot v. Estelle*, 463 U.S. 880, 926–27 (1983) (Blackmun, J., dissenting) (“Where the public holds an exaggerated opinion of the accuracy of scientific testimony, the prejudice is likely to be indelible.”); *see also infra* notes 108–09 and accompanying text (discussing “*CSI* effect”).

issue.”<sup>56</sup> In addition, “the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable,”<sup>57</sup> i.e., “trustworth[y].”<sup>58</sup> During the hearing, the party proffering the evidence must prove the technique’s reliability by a preponderance of the evidence.<sup>59</sup>

It bears reiterating that in applying the *Daubert* standard for admitting evidence under Rule 702, judges must also consider Rule 403.<sup>60</sup> The *Daubert* Court explained the conundrum of admitting expert evidence: Expert evidence is both particularly powerful and also presents great potential to be unfairly prejudicial.<sup>61</sup> While such evidence allows the jury to draw useful inferences, it is also difficult to evaluate and can mislead jurors.<sup>62</sup>

### 3. *Factors in Daubert’s Validity Inquiry*

The scientific validity inquiry—the third prong of the *Daubert* standard—is *Daubert*’s main mechanism for ensuring the quality of admitted expert evidence.<sup>63</sup> The *Daubert* Court emphasized that “[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*.”<sup>64</sup> This validity inquiry focuses on the scientific methodology used rather than the conclusions the scientist ultimately generates.<sup>65</sup> The Court discussed four flexible factors that a judge should consider in evaluating the validity of the underlying methodology: (1) testability, (2) peer review and publication, (3) error

---

<sup>56</sup> *Daubert*, 509 U.S. at 592–93. Justice Blackmun expressed faith in a judge’s capacity to evaluate the validity of different scientific methodologies. *Id.* at 593. Judges may also make this determination without conducting a *Daubert* hearing when they have enough information already to make an admissibility determination on their own. *See Oddi v. Ford Motor Co.*, 234 F.3d 136, 153–55 (3d Cir. 2000) (holding that plaintiff was not entitled to evidentiary hearing before court ruled proffered expert testimony inadmissible because there was already sufficient information to determine expert’s methodology and reliability).

<sup>57</sup> *Daubert*, 509 U.S. at 589.

<sup>58</sup> *Id.* at 590–91 n.9.

<sup>59</sup> *Id.* at 592 n.10. The *Daubert* Court found that the judge fills a gatekeeper role under Rule 104(a). *Id.* at 589; *see also* FED. R. EVID. 104(a) (“Preliminary questions concerning . . . the admissibility of evidence shall be determined by the court . . . .”); *Bourjaily v. United States*, 483 U.S. 171, 176 (1987) (“[W]hen the preliminary facts relevant to Rule 801(d)(2)(E) are disputed, the offering party must prove them by a preponderance of the evidence.”).

<sup>60</sup> *See Daubert*, 509 U.S. at 595; *see also supra* notes 30–44 and accompanying text (discussing relationship between Rules 702 and 403).

<sup>61</sup> *See Daubert*, 509 U.S. at 595. Commentators often discuss the problem of unfair prejudice. *See infra* Part II.B (discussing expert evidence’s potential for unfair prejudice).

<sup>62</sup> *Daubert*, 509 U.S. at 595.

<sup>63</sup> FAIGMAN ET AL., *supra* note 12, at 17–19.

<sup>64</sup> *Daubert*, 509 U.S. at 591 n.9.

<sup>65</sup> *Id.* at 594–95.

rate, and (4) general acceptance in the scientific community.<sup>66</sup> Each of these factors is flawed theoretically. In Part II, I show that they are particularly problematic in the criminal trial context.

The first factor, testability, examines whether a certain methodology generates hypotheses that can be tested and falsified—i.e., whether the statements and conclusions put forth by the methodology are capable of being verified.<sup>67</sup> This factor is entirely redundant when considered alongside the other three factors: If a method has an error rate, its testability is presupposed, and if a method is generally accepted or peer reviewed, it is very likely testable to begin with.<sup>68</sup> Thus, the testability prong adds little to the analysis.

For the second factor, the judge asks whether the methodology has been subject to peer review and publication. The rationale for this factor is that it “increases the likelihood that substantive flaws in methodology will be detected.”<sup>69</sup> However, the *Daubert* decision itself suggests flaws with using peer review as a factor: It does not always correlate with reliability, and sometimes “well-grounded but innovative theories will not have been published.”<sup>70</sup> This criticism is especially true if the specific methodology at issue is “too particular, too new, or of too limited interest to be published.”<sup>71</sup>

The third factor is the “known or potential rate of error.”<sup>72</sup> The Court did not provide a specific definition for rate of error.<sup>73</sup> Even theoretically, this lack of specificity is problematic for two reasons. First, while *Daubert* does say that the judge is to evaluate the underlying methodology,<sup>74</sup> the Court has not defined what “underlying methodology” means.<sup>75</sup> The decision does not specify whether error rates are to be measured at a *general* level (for the methodology

<sup>66</sup> *Id.* at 593–95.

<sup>67</sup> *Id.* at 593.

<sup>68</sup> FAIGMAN ET AL., *supra* note 12, at 20. Note that all four validity factors are not perfectly discrete: There is overlap between them to varying degrees. For instance, if a methodology is generally accepted by the scientific community, it probably has been peer reviewed. However, aside from testability, I do not deem overlap between the other factors as serious enough to have any effect on my substantive analysis. Testability is the only factor that is completely subsumed by the other factors.

<sup>69</sup> *Daubert*, 509 U.S. at 593.

<sup>70</sup> *Id.*

<sup>71</sup> *Id.*

<sup>72</sup> *Id.* at 594.

<sup>73</sup> FAIGMAN ET AL., *supra* note 12, at 26.

<sup>74</sup> *Id.* at 24.

<sup>75</sup> *Id.* at 22–24 (explaining that methodology can be defined at varying levels of specificity and that *Daubert* did not point clearly to one of those levels). *Daubert*’s attempt to demarcate cleanly between the province of the judge in evaluating the methodology and that of the jury in evaluating the conclusions from the methodology was futile: “Stated simply, the value of a methodology or of the reasoning employed in a study depends

overall, assuming perfect technique) or at the *specific* level (for the methodology as it is to be practiced according to a certain lab's particular protocols).<sup>76</sup> I will distinguish these two conceptions of the error rate as a "general error rate" and a "specific error rate."<sup>77</sup>

An example from the DNA context helps explain this distinction. The general error rate of a particular type of DNA testing is measured as the testing generally is practiced across the entire industry—i.e., it is the national error rate for that type of testing. On the other hand, the specific error rate is measured based on the specific methodology as it is practiced in the lab where the evidence was analyzed.<sup>78</sup> Particular labs tend to have "unique testing features," which are accounted for in the specific error rate.<sup>79</sup>

Second, the *Daubert* opinion does not explain whether the error rate refers to the rate of false negatives, false positives, or both.<sup>80</sup> The false positive rate measures how frequently a methodology erroneously finds a match;<sup>81</sup> the false negative rate is how often a methodology erroneously finds no match.<sup>82</sup> An error rate can also be measured by the composite of false positives and negatives—the false discovery rate.<sup>83</sup> The Court gave no guidance as to whether error-rate analysis should be limited to a single type of error rate or which type of error rate would be preferred.<sup>84</sup>

The last *Daubert* factor is general acceptance of the methodology by the scientific community, which is roughly equivalent to the *Frye*

---

entirely on what hypothesis is being studied and, thus, what conclusions are within contemplation." *Id.* at 22.

<sup>76</sup> See Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 866–75, 884 (1996) (discussing error rates at laboratory level).

<sup>77</sup> My conception of the specific error rate does not include technician error—when the lab technician deviates from the lab's protocol.

<sup>78</sup> See FAIGMAN ET AL., *supra* note 12, at 27 (discussing errors that may arise in individual labs or tests); see also Koehler, *supra* note 76, at 873–74 (discussing error rate measure that incorporates "the unique testing features associated with a particular laboratory").

<sup>79</sup> Koehler, *supra* note 76, at 871–73. Specific error rates are measured by testing labs in so-called "proficiency tests." *Id.* at 869–75.

<sup>80</sup> See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593 (1993) (discussing sources of error in DNA testing, including human error and testing features of individual labs).

<sup>81</sup> Jonathan J. Koehler, *Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter*, 59 HASTINGS L.J. 1077, 1081–82 (2008).

<sup>82</sup> *Id.* at 1082.

<sup>83</sup> *Id.* at 1083.

<sup>84</sup> See Koehler, *supra* note 81, at 1082 ("[False discovery rate is] the rate at which an examiner's claim that two markings share a common source is wrong.").

standard.<sup>85</sup> While general acceptance was the entire test in *Frye*, under *Daubert*, it is merely one factor.<sup>86</sup> General acceptance is neither a necessary nor a sufficient criterion for admission,<sup>87</sup> and its criticisms from the *Frye* context still apply.<sup>88</sup>

## II

### OUTCOMES IN CRIMINAL TRIALS UNDER THE *DAUBERT* STANDARD

Use of the *Daubert* standard in criminal trials is problematic in that it undermines two main criminal justice goals: trial accuracy and consistency.<sup>89</sup> Criminal cases have inherent adversarial biases that civil cases do not, and, in failing to account for these differences, *Daubert* causes unfair results for criminal defendants.<sup>90</sup> In addition, *Daubert*'s lack of clarity has led to inconsistent admission of expert evidence.

#### A. Resource Deficiency Problem

*Daubert* does not work well in criminal trials largely due to resource inequalities.<sup>91</sup> Because of systemic resource deficiencies, defendants' *Daubert* challenges are often perfunctory, if they take place at all.<sup>92</sup> As such, unlike in the civil litigation context, criminal defendants almost always lose their *Daubert* challenges.<sup>93</sup> Resource

---

<sup>85</sup> See FAIGMAN ET AL., *supra* note 12, at 26 (“[T]he Court did not specify what error rate should generally be expected . . .”).

<sup>86</sup> *Id.*

<sup>87</sup> *Id.*

<sup>88</sup> See *supra* note 49 and accompanying text (discussing criticisms of *Frye*).

<sup>89</sup> See *supra* Part I.A. This is not surprising given that the *Daubert* trilogy is composed of three civil, and not criminal, cases. See Neufeld, *supra* note 17, at S109 (discussing *Daubert*'s tendency to create unfair outcomes for criminal defendants).

<sup>90</sup> See generally Bernstein, *supra* note 14 (explaining *Daubert*'s failures in different types of cases, including forensic testimony in criminal litigation).

<sup>91</sup> Jack B. Weinstein, Senior U.S. Dist. Judge, E.D.N.Y., Science, and the Challenge of Expert Testimony in the Courtroom (Dec. 6, 1997), in 77 OR. L. REV. 1005, 1008 (1998) (“Courts, as gatekeepers, must be aware of how difficult it can be for some parties—particularly indigent criminal defendants—to obtain an expert to testify. The fact that one side may lack adequate resources with which to fully develop its case is a constant problem.”). The problem of adversarial bias exacerbates the resource deficiency issue. “These biases include the conscious biases of hired guns, the unconscious biases of other paid experts, and the selection biases that result from the fact that attorneys ‘shop’ for their experts from a large pool of qualified individuals.” Bernstein, *supra* note 14, at 488.

<sup>92</sup> Neufeld, *supra* note 17, at S110. One commentator has remarked that “the principal failing of *Daubert* is its misplaced reliance on a robust adversarial system to expose bad science.” *Id.* “[D]espite the frequency with which scientific and expert testimony is proffered in criminal cases, there is a dearth of *Daubert* challenges and hearings.” *Id.* at S109.

<sup>93</sup> *Id.* at S109. On the other hand, prosecutors win almost all of their *Daubert* challenges. *Id.*

deficiency is problematic because it will often undermine defense lawyers' ability to challenge the prosecution's experts.<sup>94</sup> Ideally, if an expert gives misleading testimony, the weakness should be exposed through cross-examination. However, the defense usually does not have the resources or knowledge base necessary to contest faulty expert testimony adequately.<sup>95</sup> As a result, the prosecution may be able to admit misleading forensic evidence,<sup>96</sup> an ability which translates into a higher likelihood of false convictions.<sup>97</sup>

*Daubert* fails to account for this inequality of resources. While faulty admission standards for expert evidence are not responsible for all false convictions, these standards likely account for a nontrivial portion of such convictions.<sup>98</sup> This phenomenon is deeply troubling given one recent study that found a 3.3% wrongful conviction rate in capital rape-murders.<sup>99</sup>

### B. *Unfair Prejudice Problem with Expert Evidence*

Admitting forensic evidence of dubious validity is particularly problematic in the jury trial context because such evidence can often be unfairly prejudicial, thereby exacerbating the false conviction problem.

First, juries are, as a general matter, unskilled at interpreting statistical evidence and testimony.<sup>100</sup> They are notoriously bad at assessing the validity of scientific evidence and have particular diffi-

---

<sup>94</sup> DNA exonerations often demonstrate that wrongful convictions stem from material misstatements of fact by forensic scientists. The most common problems that arise include conclusions without any scientific basis, reports that ignore or deliberately distort data, testimony that ignores or deliberately distorts the report, and testimony and report writing clearly beyond the competence of the examiner. *Id.* at S110.

<sup>95</sup> Garrett & Neufeld, *supra* note 7, at 33.

<sup>96</sup> *Id.* at 34.

<sup>97</sup> While there is no one-to-one causality here, more misleading evidence against criminal defendants generally correlates to more false convictions. *See id.* at 4–5 (showing how invalid forensic testimony can lead to false conviction).

<sup>98</sup> We can infer the contribution of improper admission standards to false convictions from the relationship between faulty forensic testimony and the number of false convictions. *See generally* Garrett & Neufeld, *supra* note 7 (exploring relationship between erroneous forensic testimony and wrongful convictions).

<sup>99</sup> D. Michael Risinger, *Innocents Convicted: An Empirically Justified Wrongful Conviction Rate*, 97 J. CRIM. L. & CRIMINOLOGY 761, 778–80 (2007). Overall, “estimates of the rate of wrongful convictions range between a low of 0.5% to a high of 15% with between 1% and 3% being the estimate with the most empirical support.” Gia E. Barboza & Lynn M. Goedecke, *Can Federal Oversight of Forensic Science Reduce the Incidence of Wrongful Convictions?*, 45 CRIM. L. BULL. 881, 881 (2009).

<sup>100</sup> Much empirical work documents this problem. *E.g.*, Bradley D. McAuliff et al., *Can Jurors Recognize Missing Control Groups, Confounds, and Experimenter Bias in Psychological Science?*, 33 LAW & HUM. BEHAV. 247 (2009) (evaluating ability of jurors to measure internal validity threats in psychological scientific testimony).

culty weighing probabilities and understanding statistics generally.<sup>101</sup> There is evidence that juries frequently overestimate the significance of a forensic match between a defendant's sample and the police sample.<sup>102</sup> For instance, juries misconstrue DNA evidence by equating the probability of a DNA match with a probability of guilt (when it is often just the probability that the defendant was present at the scene).<sup>103</sup> This inability to weigh probabilities accurately can cause forensic evidence to be unfairly prejudicial.<sup>104</sup>

Second, juries can overvalue expert evidence and fail to fully evaluate its quality.<sup>105</sup> There is considerable evidence that juries trust expert testimony implicitly.<sup>106</sup> In his famed *Barefoot v. Estelle* dissent, Justice Blackmun wrote, "Where the public holds an exaggerated opinion of the accuracy of scientific testimony, the prejudice is likely to be indelible."<sup>107</sup> In the context of forensic science, this phenomenon has been called the "CSI effect," and is based on the assumption that jurors hold an unrealistically high expectation that forensic evidence will be presented at trial.<sup>108</sup> According to some commentators,

<sup>101</sup> See *id.* at 248 (citing several studies indicating that mock jurors have same difficulties evaluating statistical and methodological information as lay public at large).

<sup>102</sup> Tom R. Tyler, *Viewing CSI and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction*, 115 *YALE L.J.* 1050, 1068 (2006) ("There is widespread evidence indicating that people . . . overestimate the probative value of scientific evidence."). One example is the "source probability error" in the DNA evidence context, which describes situations when the jury mistakenly believes that the match probability given by an expert is the same as the probability that someone other than defendant left his sample at the crime scene. Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 *JURIMETRICS J.* 21, 27 (1993). In other words, if the expert says the probability of a match between crime scene and defendant DNA samples is thirty-three percent, the jury mistakenly believes that thirty-three percent is the probability that someone other than the defendant left their DNA at the crime scene.

<sup>103</sup> See Koehler, *supra* note 102, at 31–32 (describing "ultimate issue error").

<sup>104</sup> See *United States v. Shea*, 957 F. Supp. 331, 344 (D.N.H. 1997) (acknowledging that because probabilities in DNA profile matching have potential to mislead jurors, probative value and unfair prejudicial value of evidence must be balanced), *aff'd*, 159 F.3d 37 (1st Cir. 1998).

<sup>105</sup> See Giannelli, *supra* note 49, at 1237 (noting tendency of scientific evidence to mislead juries).

<sup>106</sup> See *id.* ("[A]n aura of scientific infallibility may shroud the evidence and thus lead the jury to accept it without critical scrutiny."); Tyler, *supra* note 102, at 1068 ("There is widespread evidence indicating that people . . . overestimate the probative value of scientific evidence.").

<sup>107</sup> 463 U.S. 880, 926–27 (1983) (Blackmun, J., dissenting). While acknowledging that expert psychologists "are accurate in no more than one out of three predictions of violent behavior," the majority nevertheless expressed confidence in the adversarial system's ability to "sort out the reliable from the unreliable evidence and opinion." *Id.* at 899–901 & n.7.

<sup>108</sup> N.J. Schweitzer & Michael J. Saks, *The CSI Effect: Popular Fiction About Forensic Science Affects the Public's Expectations About Real Forensic Science*, 47 *JURIMETRICS J.* 357, 358 (2007); see also Craig M. Cooley, *The CSI Effect: Its Impact and Potential Con-*

modern-day forensics crime shows have led people to believe that “crime labs are pristine scientific sanctuaries” and that forensic scientists make few, if any, mistakes.<sup>109</sup>

Because resource inequalities lead to the admission of evidence of dubious validity, it is particularly important that juries have the ability to ascertain when forensic evidence is of low quality. Unfortunately, bad evidence often sways juries, thwarting the first criminal justice goal of reducing false convictions.<sup>110</sup>

### C. Consistency and Workability Problems for Judges

The excessive flexibility of the *Daubert* validity test fails to promote the criminal justice goal of consistency and predictability in trial outcomes.

While the *Daubert* Court emphasized that this inquiry was a flexible one,<sup>111</sup> in reality, many lower court judges have interpreted the *Daubert* factors as a “mechanical checklist.”<sup>112</sup> And while many courts do use the *Daubert* factors as a flexible test, they vary greatly in the degree of flexibility that they actually employ, with some courts using the factors too loosely.<sup>113</sup> Furthermore, because the standard of review for *Daubert* admissions decisions is abuse of discretion,<sup>114</sup> trial judges’ applications of this test are effectively unreviewable.<sup>115</sup>

---

*cerns*, 41 NEW ENG. L. REV. 471, 471–73 (2007) (describing distorted perceptions of viewers of forensic science crime dramas). *But see* Tyler, *supra* note 102, at 1076–83 (expressing skepticism that “*CSI* effect” exists).

<sup>109</sup> Cooley, *supra* note 108, at 471. Commentators disagree over whether such an effect exists, and, if it does, whether it hurts the prosecution or the defense. *See generally* Schweitzer & Saks, *supra* note 108. On the one hand, the *CSI* effect could mean that jurors need more forensic science to persuade them of guilt, which helps the defense. Alternatively, the *CSI* effect could mean that jurors blindly believe any forensic evidence presented by the prosecution. *Id.* at 358. “[T]he commentators seem to agree on one thing: that *CSI* is convincing the public that forensic science not only is science, but it is super science.” *Id.*

<sup>110</sup> *See* Garrett & Neufeld, *supra* note 7, at 5 (discussing faulty work in crime laboratories leading to post-conviction exonerations).

<sup>111</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 594 (1993).

<sup>112</sup> D. Michael Risinger, *Goodbye to All That, or a Fool's Errand, by One of the Fools: How I Stopped Worrying About Court Responses to Handwriting Identification (and "Forensic Science" in General) and Learned To Love Misinterpretations of Kumho Tire v. Carmichael*, 43 TULSA L. REV. 447, 460 (2007).

<sup>113</sup> *See* Welch, *supra* note 16, at 1091 (describing flexible application of *Daubert* factors by some courts as problematic). Further, some courts seem to register no real understanding of the *Daubert* validity factors at all. *See* David L. Faigman, *Admissibility Regimes: The "Opinion Rule" and Other Oddities and Exceptions to Scientific Evidence, the Scientific Revolution, and Common Sense*, 36 SW. U. L. REV. 699, 718–22 (2008) (noting that some courts apply *Daubert* factors “in a fantastically ignorant fashion”).

<sup>114</sup> *See* Welch, *supra* note 16, at 1092–94 (noting negative effects of abuse of discretion review of trial court *Daubert* admissions).

<sup>115</sup> *Id.* at 1094.

When judges use the *Daubert* factors too flexibly, they are not acting as gatekeepers in policing the admission of expert evidence. Overflexibility leads to admission of “junk science”—evidence of poor probative value that can also be highly misleading to jurors.<sup>116</sup> Thus, the flexibility in *Daubert*’s validity inquiry leads to unpredictability and inaccuracy in trial outcomes, thwarting both criminal justice goals.

The *Daubert* Court’s deficient explanation of the error rate factor further exacerbates these problems. Since judges have little guidance as to how to measure a relevant error rate and its acceptable magnitude, the validity inquiry under *Daubert* does not function as it should to catch unreliable evidence before it reaches the jury.<sup>117</sup>

### III

#### A DOCTRINAL SOLUTION: TWEAKING THE *DAUBERT* STANDARD TO BETTER SERVE CRIMINAL JUSTICE GOALS

It is clear that *Daubert* is failing to promote criminal justice goals. Many have proposed solutions, often regulatory or administrative.<sup>118</sup> I propose a doctrinal tweak that would reformulate the case law interpreting Rules 702 and 403 for forensic evidence proffers in criminal trials. First, in light of the aforementioned problems with forensic evidence in criminal trials, I argue that there should be a rebuttable presumption against the admission of expert evidence. The *Daubert* validity inquiry should also be restructured to promote the two criminal justice goals: The specific error rate should be the “first” factor in the *Daubert* validity inquiry, with other factors to be examined *if and only if* the specific error rate cannot be ascertained. Additionally, I propose two ways to develop a more coherent and concrete definition of the error rate factor, both of which hinge on the type of error present in the evidence proffer.

This doctrinal tweak would better serve the goals of the criminal justice system. My proposal more accurately reflects Rule 403’s concern with unfair prejudice<sup>119</sup> and fills the resource inequality lacuna that allows for the admission of evidence of dubious validity. As a result, juries would reach more accurate results, with fewer false con-

---

<sup>116</sup> *Id.*

<sup>117</sup> The lack of guidance on the error rate prong is exemplified by the notorious case *United States v. Havvard*, 117 F. Supp. 2d 848 (S.D. Ind. 2000), in which the court accepted the government’s contention that the error rate of latent fingerprint identification is zero. *Id.* at 854. At the very least, given the possibility of human error, no forensic method could have an error rate of zero. Such blunders reveal a misunderstanding of the basic conception of an error rate.

<sup>118</sup> See *supra* notes 17–18 (listing regulatory and administrative proposals).

<sup>119</sup> See *supra* Part I.B.2 (discussing Rule 403).

victions. In addition, this reformulation lends more structure and clarity to the *Daubert* validity inquiry, promoting the second goal of trial outcome predictability.<sup>120</sup>

### A. *Reversing the General Presumption of Admitting Expert Evidence*

To promote trial outcome accuracy,<sup>121</sup> my doctrinal reformulation starts with a shift that is similar to general suggestions for expert evidence admissibility proposed by others: Force parties introducing forensic evidence to overcome a rebuttable presumption that their evidence is unduly prejudicial and thus inadmissible.<sup>122</sup>

Most evidence proffers benefit from the presumption of admissibility embodied in Rule 403.<sup>123</sup> However, this presumption should be reversed for forensic evidence proffers in criminal trials. *Daubert's* shortcomings with regard to criminal justice goals justifies this presumption's reversal.<sup>124</sup> First, forensic evidence has a particularly high potential to be unfairly prejudicial.<sup>125</sup> Moreover, this reversal is hardly unprecedented; other Federal Rules of Evidence reverse the Rule 403 presumption of admissibility for particularly unfair prejudicial evidence, creating a presumption *against* admissibility.<sup>126</sup>

---

<sup>120</sup> It is true that this approach is more restrictive with respect to admitting expert evidence and runs against *Daubert's* original philosophy of establishing a more flexible admissions standard. See *Daubert*, 509 U.S. 579, 594 (1993). However, given the grave shortcomings of *Daubert* in the criminal context described above, and the fact that the *Daubert* trilogy explicated the doctrine in the civil litigation context only, a more restrictive standard is warranted in criminal trials.

<sup>121</sup> See *supra* Part I.A (discussing criminal justice goals).

<sup>122</sup> For example, Professor Murphy suggests that "reliability [should be] treated as a threshold question of admissibility, on which the proponent of the evidence carry[s] the burden, [and] then the government would have to submit evidence of the laboratory's error rate." Murphy, *supra* note 18, at 796. In fact, the joint symposia of the National Conference of Lawyers and Scientists and the ABA Section of Science and Technology in the 1980s discussed instituting such a presumption against scientific evidence admissibility into Rule 702. Kenneth Kreiling, *Managing Expert Evidence: An Overview*, 36 JURIMETRICS J. 121, 122 & n.7 (1996) (reviewing FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (1994)).

<sup>123</sup> See *supra* note 43 and accompanying text (discussing presumption of admissibility in Rule 403).

<sup>124</sup> See *supra* Part II.A–B (discussing *Daubert's* failings in criminal trials).

<sup>125</sup> See *supra* Part II.B (explaining potential of forensic evidence to be unfairly prejudicial).

<sup>126</sup> For example, the "direct connection doctrine" governing third-party perpetrator evidence reverses the general Rule 403 presumption of admissibility. Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 WIS. L. REV. 291, 360 (2006). A presumption reversal also exists in the Federal Rules. Evidence of conviction of a crime "punishable by death or imprisonment in excess of one year" is admissible for impeachment of a witness "if the court determines that the probative value of admitting this evidence outweighs its prejudicial effect to the accused." FED. R. EVID.

The prior bad acts doctrine is a key example of this presumption against admissibility.<sup>127</sup> Evidence of prior bad acts is so likely to be unfairly prejudicial that it presumptively fails the Rule 403 balancing test.<sup>128</sup> Rule 404(b)'s presumption against admitting such evidence to demonstrate criminal propensity reflects this likelihood of prejudice.<sup>129</sup> Though this evidence can be highly probative, it is too likely that the jury would weigh it improperly.<sup>130</sup> Similarly, juries on the whole weigh forensic evidence improperly, a fact which justifies a presumption against admitting such evidence.<sup>131</sup>

The adversarial bias problem further justifies a presumption against admitting forensic evidence. Parties in litigation face imperfect information about the practices of the crime labs generating their opponents' proffers of expert evidence.<sup>132</sup> In particular, defendants have a harder time than the prosecution in accessing information about the police crime lab and its methods.<sup>133</sup> As a result, it is more costly for defendants' lawyers to challenge prosecution proffers competently.

The proffesor should rebut the presumption against expert evidence admissibility by showing that the probative value of the evi-

---

609(a). The burden is on the government to prove that the evidence should be admitted, a requirement which reverses the Rule 403 presumption. Jeffrey Bellin, *Circumventing Congress: How the Federal Courts Opened the Door to Impeaching Criminal Defendants with Prior Convictions*, 42 U.C. DAVIS L. REV. 289, 309 & n.77 (2008).

<sup>127</sup> See FED. R. EVID. 404(b) (codifying common law prior bad acts doctrine); *Old Chief v. United States*, 519 U.S. 172, 181–82 (1997) (describing common law presumption against admission of evidence to discredit defendant's character); *Michelson v. United States*, 335 U.S. 469, 475–76 (1948) (same).

<sup>128</sup> See *Michelson*, 335 U.S. at 475–76 (noting that evidence of prior bad acts “is said to weigh too much with the jury and to so overpersuade them as to prejudice one with a bad general record”).

<sup>129</sup> See FED. R. EVID. 404(b) (“Evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith.”). This rule codified the common law prior bad acts doctrine. See *Old Chief*, 519 U.S. at 181 (explaining that Rule 404(b) “reflects th[e] common-law tradition” of disallowing propensity evidence). However, evidence of prior bad acts may come in for certain enumerated purposes including “proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident,” if the prosecution gives reasonable notice. FED. R. EVID. 404(b).

<sup>130</sup> See *Old Chief*, 519 U.S. at 181 (“Although . . . ‘propensity evidence’ is relevant, the risk that a jury will convict for crimes other than those charged—or that, uncertain of guilt, it will convict anyway because a bad person deserves punishment—creates a prejudicial effect that outweighs ordinary relevance.” (quoting *United States v. Moccia*, 681 F.2d 61, 63 (1st Cir. 1982))).

<sup>131</sup> See *supra* Part II.B (explaining juries' lack of skill in interpreting and assessing validity of statistical and scientific evidence).

<sup>132</sup> See *Murphy*, *supra* note 18, at 790–91 (emphasizing importance of defendant's receiving access to government's scientific and statistical information).

<sup>133</sup> *Id.*

dence outweighs its prejudicial value.<sup>134</sup> This proposal requires a higher burden than preponderance of evidence—such as a clear and convincing evidence standard<sup>135</sup>—that *Daubert* is satisfied. This *Daubert* showing would also include my proposed validity inquiry into the specific error rate.<sup>136</sup> In the absence of error rate information, the “secondary” validity factors of peer review and publication and general acceptance could be used to rebut the presumption.

Finally, I tentatively suggest that such a presumption apply only to prosecutorial proffers of forensic evidence. I propose this limitation due to a concern that such a presumption could otherwise prevent indigent defendants from introducing exculpatory forensic evidence. A detailed defense of this position, however, is beyond the scope of this Note.<sup>137</sup>

---

<sup>134</sup> This approach is a direct reversal of the Rule 403 presumption in favor of admissibility and is how the presumption against admissibility is rebutted in the Rule 609(a)(1) context. See FED. R. EVID. 609(a)(1) (governing prior convictions of accused); *United States v. Sanders*, 964 F.2d 295 (4th Cir. 1992); see also *supra* note 126 (explaining reversal of presumption in Rule 403 context). The test to determine whether the Rule 609(a)(1) presumption has been rebutted is from *Gordon v. United States*, 383 F.2d 936, 939–41 (D.C. Cir. 1967).

<sup>135</sup> Detailed consideration of the burden on the proffering party is beyond the scope of this Note. What is clear, however, is that the showing of reliability must be higher than the preponderance of the evidence standard required in the current *Daubert* test. See *supra* note 59 and accompanying text (noting that party proffering evidence must prove technique’s reliability by preponderance of evidence).

<sup>136</sup> The onus is on the party introducing the evidence to produce enough information to satisfy the judge that she has relevant specific error rate information. Cf. *Murphy*, *supra* note 18, at 793 (“Accordingly, the government should carry a burden of placing before the court continued evidence of a technique’s legitimacy.”). I decline to prescribe a threshold error rate. See *infra* notes 163–64 and accompanying text (explaining difficulties in setting threshold error rate).

<sup>137</sup> A burden-shifting scheme makes it difficult for the defense as well as the government to introduce forensic evidence. Thus, there is a potential conflict with the first criminal justice goal of mitigating false convictions. The defense may not have the resources or access to gather error rate information, circumstances which could potentially result in more harm than good for defendants.

There are two possible responses to this predicament. First, one could say that the additional burden on the defense does not matter—the vast majority of forensic evidence is introduced by the prosecution regardless of the burden, so there is a large net benefit to the defense from such a scheme. See *Garrett & Neufeld*, *supra* note 7, at 90 (“The presentation of forensic science testimony is typically one-sided in the majority of states that do not routinely fund the provision of forensic experts for indigent defendants.”).

Second, the rebuttable presumption could apply only when the prosecution seeks to introduce evidence against a defendant as is the case elsewhere in the Rules of Evidence. Rules 404(b) and 609(a)(1) are ready examples. See *supra* notes 126, 129, 134 and accompanying text (discussing reversal of presumption of admissibility for particularly unfair prejudicial evidence).

Because of costs to individual defendants, I prefer the second response. Whichever response is chosen, a presumption against forensic evidence admission has large net benefits to defendants in curing resource inequality and unfair prejudice problems.

## B. *Setting the Error Rate as the Primary Factor in Daubert's Validity Inquiry*

In order to better serve criminal justice goals,<sup>138</sup> the *Daubert* validity inquiry should focus on the specific error rate—the error rate of the methodology as practiced in that particular instance.<sup>139</sup> Publication, general acceptance, and the general error rate should be considered as secondary factors that are relevant if and only if the specific error rate cannot be ascertained at all.<sup>140</sup>

### 1. *Setting the "Specific" Error Rate as the Primary Validity Factor*

The error rate factor under the validity inquiry of the *Daubert* standard is the single most important factor that reflects the probative value of expert evidence. Though the *Daubert* Court never explicitly defined the term, “[t]he typical use of the term ‘error rate’ refers to the number of ‘mistakes’ a particular technique or method will make in some specific number of trials.”<sup>141</sup> Validity is measured most directly by the error rate: When we are concerned about accuracy, we want to know how often the evidence is wrong.<sup>142</sup>

Prioritizing the error rate also maximizes the probative value of evidence.<sup>143</sup> The lower the error rate of a scientific methodology, the more likely the methodology is accurate, and thus the more likely that it provides probative evidence.<sup>144</sup> The error rate is therefore a concrete and objective way to measure the probative value of evidence.

Take the example of bite mark evidence. Imagine Crime Lab X uses a procedure that largely resembles standard bite mark testing. However, the compound Lab X uses to capture the defendant’s bite impression is unique and used infrequently by other labs. When evaluating the underlying methodology of the bite mark evidence, the ultimate question is how often Lab X’s unique testing procedure yields inaccurate conclusions. Knowing how often this novel testing method

<sup>138</sup> See *supra* Part I.A (describing criminal justice goals).

<sup>139</sup> See *supra* notes 72–79 and accompanying text (explaining concept of specific error rate).

<sup>140</sup> The last factor, testability, is presupposed by the other three factors, and I therefore do not consider it in my proposed reformulation. FAIGMAN ET AL., *supra* note 12, at 20.

<sup>141</sup> *Id.* at 25.

<sup>142</sup> Other commentators have presupposed that the error rate is the most important factor and the best measure of reliability. See, e.g., Murphy, *supra* note 18, at 795–96 (advocating for threshold error rate for admissibility); Barry C. Scheck, *DNA and Daubert*, 15 CARDOZO L. REV. 1959, 1997 (1994) (“*Daubert* analysis of DNA evidence reveals that laboratory error rate is the most important reliability factor for courts to consider.”).

<sup>143</sup> See *supra* Part I.B.2 (discussing Rule 403 balancing).

<sup>144</sup> See Koehler, *supra* note 76, at 866 (drawing link between error rate of DNA testing method and probative value of evidence in O.J. Simpson case).

produces bad results allows a true assessment of the value of the evidence.

This example reveals how to define the *Daubert* error rate: Instead of being measured at a general methodological level, it should be measured at a lab-specific methodological level.<sup>145</sup> In short, the “error rate” should mean the *specific* error rate.<sup>146</sup> A specific error rate measure provides more relevant information for a particular adjudication because it accounts for additional details of the relevant testing conditions.<sup>147</sup> It encompasses the lab’s particular protocols as to how it carries out a forensic testing methodology and it takes into account how the lab’s technicians perform that methodology.<sup>148</sup> “Proficiency testing”—under which a lab performs the methodology in question in controlled trials—can be used to test and record the methodology’s specific error rate.<sup>149</sup>

The specific error rate provides a more accurate reflection of the probative value of a piece of evidence in a specific trial than does the general error rate. Crucially, if the error rate were based on a broader definition of the underlying methodology, juries would have to ascertain how closely a lab’s practices in that particular instance comport with the general methodology for which the jury has general error rate information—a common situation in the current jurisprudence.<sup>150</sup>

Two of *Daubert*’s weaknesses in the criminal context are also implicated when using the general error rate. Given jurors’ lack of expertise and inability to weigh expert evidence properly, it is unwise to trust them to make sound judgments about the error rate.<sup>151</sup> The

---

<sup>145</sup> See *supra* notes 74–85 and accompanying text (explaining concept of specific error rate).

<sup>146</sup> This idea raises the question of just how specific this error rate should be. For example, should it be the error rate for each individual technician? Error rates for individual technicians are likely too specific to be practicable. Instead, the error rate should be measured for that particular type of forensic testing across all of the lab’s technicians. This is specific enough to measure a lab’s error rate overall while also being a practicable way to measure such an error rate.

<sup>147</sup> See Koehler, *supra* note 76, at 870–75 (noting preference for lab-specific rather than industry-wide error data when sufficient lab-specific data is available).

<sup>148</sup> *Id.* (comparing viewpoints on use of industry-wide error rates in making error predictions for specific cases); Murphy, *supra* note 18, at 759–60 (discussing pros and cons of courts’ “case-specific” inquiries into proffered forensic evidence methodologies).

<sup>149</sup> Although some labs already perform proficiency testing, commentators such as Jonathan Koehler seem to advocate for more widespread proficiency testing of crime labs. See Koehler, *supra* note 76, at 874–75 (discussing proficiency testing’s benefits with regard to promulgating information about error rate).

<sup>150</sup> Murphy, *supra* note 18, at 758–59.

<sup>151</sup> See McAuliff et al., *supra* note 100, at 248 (demonstrating juries’ difficulties evaluating statistics in expert testimony). My conception of the specific error rate retains the jury’s role as decider of issues of fact. The specific error rate is still cabined entirely to the *methodology* that the lab is supposed to perform and can thus be comfortably character-

resource inequality problem perpetuates jurors' inability to evaluate the general error rate: When defendants are unable to raise worthy challenges to the prosecution's experts, jurors will not have access to comprehensive information with which to evaluate how the scientific testing was performed.<sup>152</sup> As such, the underlying methodology that the *Daubert* error rate measures should be defined as the specific error rate.

A presumption against the admission of forensic evidence would give prosecutors' offices and police departments an incentive to maintain complete information about their crime labs' methodologies and specific error rates.<sup>153</sup> This incentive significantly bolsters the practicality and effectiveness of my proposed *Daubert* validity inquiry: Even if specific error rates for labs are not initially available, the presumption will encourage labs to make specific information about the error rate known more readily.<sup>154</sup> Furthermore, since the burden is on the party introducing the evidence, the costs of producing such information are on the parties and do not have to be swallowed by the court in hiring its own experts to ascertain the specific error rate.

The specific error rate is the most objective and rigorous way to measure the validity of a forensic methodology, especially compared to the other two substantial validity factors under *Daubert*: publication in a peer-reviewed journal and general acceptance of the methodology by the scientific community. These factors can provide useful information, but they suffer several shortcomings that the specific error rate is not subject to.

---

ized as an issue of law that the judge can decide without invading the province of the jury. The jury must still decide whether in a given instance the lab technician performed the testing in accordance with the articulated methodology. The tendency to overvalue expert evidence may still exist, as juries may inaccurately evaluate the competence of a specific lab technician in the current instance. FAIGMAN ET AL., *supra* note 12, at 23–24; *cf.* Koehler, *supra* note 76, at 870–75 (advocating for lab proficiency testing).

<sup>152</sup> See Murphy, *supra* note 18, at 760–64, 768 n.204 (explaining difficulties defendants face in gaining information about forensic techniques and need for jurors to receive such information); Neufeld, *supra* note 17, at S110 (explaining defense lawyers' lack of skill and funding for mounting *Daubert* challenges).

<sup>153</sup> See Murphy, *supra* note 18, at 795–96 (noting that, if courts viewed “error rates of either the methodology itself or the executing laboratory” as “legal questions of admissibility,” laboratories would have more incentive to “document their rate of error”). They would achieve this goal by instituting performance testing of their labs. See *supra* note 149 and accompanying text.

<sup>154</sup> See Murphy, *supra* note 18, at 796 (arguing that courts' determining admissibility of forensic evidence by reviewing labs' general reliability would encourage labs to better comply with operational standards and make error rate information widely available).

Publication in peer-reviewed journals and general acceptance are often skewed toward methods that are already well established.<sup>155</sup> Such acceptance may have more to do with ease of method than accuracy of results.<sup>156</sup> Also, both the publication and general acceptance factors reflect the standards of scientific communities as to the acceptable level of predictive power of certain scientific methodologies; however, since courts' expectations for the predictive power of forensic evidence introduced in litigation may be more stringent than that of scientific communities, using peer review and general acceptance factors may not be appropriate.<sup>157</sup> Further, it may be difficult to ascertain the general practice of a scientific community; indeed, no general practice may even exist.<sup>158</sup> It is clear that the specific error rate is superior to these other *Daubert* factors as a measure of reliability because peer review and general acceptance remain imperfect proxies for the value of expert evidence.

Since the specific error rate is such an important factor in its objective measure of the quality of evidence, when possible it should be the *only* factor under consideration. This approach would help promote both criminal justice goals. First, the specific error rate mathematically reflects the probative value of expert evidence, a characteristic which promotes accuracy in trial outcomes. Second, applying a single factor—the error rate—is much easier than applying the amorphous and flexible multifactor jumble that the *Daubert* standard requires today. This ease of application would provide consistency and accuracy across trial outcomes.

Finally, how low an error rate should be to rebut the presumption of admissibility remains an open question. In one way, it would be ideal to be able to articulate a threshold error rate: It would be the easiest possible validity standard judges could apply, promoting con-

---

<sup>155</sup> FAIGMAN ET AL., *supra* note 12, at 27–28 (noting mainstream journals' bias in favor of "conventional scholarship").

<sup>156</sup> Tara Marie La Morte, Comment, *Sleeping Gatekeepers: United States v. Llera Plaza and the Unreliability of Forensic Fingerprinting Evidence Under Daubert*, 14 ALB. L.J. SCI. & TECH. 171, 187 (2003) (explaining that forensic scientists "do not expend much effort researching," which allows lower-quality methods to pass general acceptance and peer review inquiries).

<sup>157</sup> See Thomas M. Fleming, Annotation, *Admissibility of DNA Identification Evidence*, 84 A.L.R.4th 313, 327 (1991) ("[T]esting laboratories have sometimes used stricter match criteria for determining band frequencies within the data base than they have used for determining a match between prints obtained from the samples in issue, resulting in an inordinately low coincidental match probability."). Even though many crime labs may use this criteria-setting method, it should be a practice that raises red flags as to the evidence it produces for trial.

<sup>158</sup> Giannelli, *supra* note 49, at 1215–19 (discussing problems with each of courts' three recognized methods of determining general acceptance).

sistency in the quality of evidence admitted and trial outcome predictability. However, I adopt the view that, because the cost of error varies in different contexts and proceedings, an across-the-board threshold is unwise.<sup>159</sup> Even if I were to prefer setting a threshold error rate, prescribing one is an extremely complicated endeavor beyond the scope of this Note.<sup>160</sup>

## 2. *Keeping General Acceptance and Peer Review as Secondary Factors*

The peer review and publication and general acceptance factors should not be discarded altogether; instead, they should be used when the error rate cannot be ascertained. Reserving the ability to use these factors solves a problem that necessarily arises if the error rate is defined as the specific error rate, because this information is often unavailable.<sup>161</sup> However, sometimes courts do have information that a crime lab's specific error rate is high and will refuse to admit evidence generated by that lab.<sup>162</sup> Sometimes the lab's practices will comport with the general methodology as it is ideally practiced—in those cases, the general error rate would apply because it would sufficiently approximate the specific error rate.<sup>163</sup>

However, for situations where the judge lacks the appropriate specific error rate information, general acceptance and peer review and publication of the specific testing method may be useful factors. It

---

<sup>159</sup> See FAIGMAN ET AL., *supra* note 12, at 26 (arguing that effects of error in given context should guide court's evaluation). In other words, depending on the realities of different adjudications, different error rates may be appropriate. For example, a lower error rate may be appropriate in capital cases as opposed to in probation matters. *Id.*

<sup>160</sup> Some commentators have advocated for the institution of a threshold specific error rate for the admission of expert evidence. *E.g.*, Murphy, *supra* note 18, at 796. Such a threshold would have to comport with the evidentiary standard for criminal conviction—proof beyond a reasonable doubt. This is a fascinating subject, but beyond the scope of this Note.

<sup>161</sup> Scheck, *supra* note 142, at 1997 (“Unfortunately, forensic laboratories have historically resisted external blind proficiency testing and other efforts to assess laboratory error rates.”).

<sup>162</sup> *E.g.*, Murray v. State, 838 So. 2d 1073, 1080–81 (Fla. 2002) (denying admission of DNA evidence in part on “general sloppiness” of methods used by lab). Such information becomes available through proficiency testing. Koehler, *supra* note 76, 870–75 (discussing arguments for and against using error rates in determining probative value of DNA evidence).

<sup>163</sup> Judges will thus retain some discretion because they will have the authority to determine whether a lab's specific practices are the same as general best practices in order for them to have appropriate error rate information. Some may argue that a judge would have too much discretion on an issue of *fact*—ascertaining whether the lab's practices are the same as best practices so that he can apply the general error rate.

is true that these factors are problematic.<sup>164</sup> Nonetheless, they are not so problematic that no useful information about the validity of a scientific methodology can ever be gleaned from them: These factors are simply *less* useful than the specific error rate factor. So in these cases alone, judges may opt to look to general practice or publication to see if the scientific community approves of the methodology used in that case.

This reformulation works because using general practice and publication as backup factors still incentivizes labs to use proficiency testing to develop specific error rate information.<sup>165</sup> There are limited instances now in which lab-specific error rate information is available,<sup>166</sup> but prosecutors will have an increased incentive to develop it if confronted with this revised doctrine.<sup>167</sup> It would be significantly easier and less risky for prosecutors to introduce forensic evidence if they knew the specific error rate. Since general practice and publication are less concrete factors than the error rate,<sup>168</sup> how judges will act on them is also less predictable.

This proposal emphasizes using the error rate to measure the probative value of evidence, an approach which satisfies the first criminal justice goal of trial accuracy. A clear standard utilizing the specific error rate as the first and, when possible, the only validity factor should promote the goal of trial outcome consistency as well.

### C. *Defining the Error Rate: Type I or Type II?*

While it should be the most important *Daubert* validity factor, the error rate is unfortunately the least clear. The *Daubert* Court never explained whether the error rate means false negatives, false positives, or the false discovery rate.<sup>169</sup> There are two possible definitions for the *Daubert* error rate that best meet the two criminal justice goals.

---

<sup>164</sup> See *supra* notes 155–160 and accompanying text (discussing varying concerns with factors).

<sup>165</sup> See Murphy, *supra* note 18, at 759 (noting appeal and ease of courts' reliance on previous findings of methodological reliability, specifically those contained in scientific journals and papers).

<sup>166</sup> See Koehler, *supra* note 76, at 874 (recognizing small number of laboratories participating in high quality proficiency tests).

<sup>167</sup> Murphy, *supra* note 18, at 796 (“[S]taking admissibility of evidence on a laboratory’s general reliability not only creates incentives on the part of the laboratory to comply with published standards of operation, but it also gives the prosecution a vested interest in competently managing the laboratory . . .”).

<sup>168</sup> See *supra* notes 155–160 and accompanying text (discussing shortcomings of these factors).

<sup>169</sup> Risinger, *supra* note 112, at 461 (“[C]asual invocation of these [varying notions of error rate] by the Supreme Court was unlikely to provide much guidance to the lower court[s] . . .”).

One ensures a high degree of trial outcome accuracy—with a particularly high emphasis on preventing false convictions—and features a high level of clarity that would promote trial outcome consistency. The second ensures a higher level of trial outcome accuracy in general but a lesser degree of consistency.

### 1. *Type I and Type II Errors*

Under the Court's decision in *Daubert*, it is unclear whether "error rate" refers to false positives or false negatives.<sup>170</sup> For clarity, I will refer to these types of error in statistics jargon: Type I errors are false positives, and Type II errors are false negatives.<sup>171</sup>

In order to answer whether *Daubert's* error rate should be Type I, Type II, or both, one should consider which type of error is implicated more frequently, rather than what the error rate of the practice is in the abstract when removed from the adjudication context. Thus, it makes sense to map out the possible errors and focus on their likelihood.

Below is a two-by-two matrix breaking down the possible error types that may be implicated in the four different situations that arise in criminal cases. The example methodology used is handwriting analysis. In each of these four situations, there is either a higher likelihood of false convictions or false acquittals.<sup>172</sup> Whether the probability of acquittals or convictions increases depends on two factors: first, who proffers a piece of expert evidence, and second, which type of error is likely in that proffer of evidence.

Since parties obviously will not proffer evidence that does not benefit their case, the identity of the party offering the evidence predicts the outcome of its scientific testing. Prosecutors will proffer

---

<sup>170</sup> *Id.*

<sup>171</sup> Courts often define the error rate as Type I when they accept evidence based on hypothesis testing. *E.g.* *Love v. Johans*, 439 F.3d 723, 731 (2006) (inquiring why multiple regression analysis was not submitted). Some common forensic methodologies, such as fingerprint analysis, are not based on hypothesis testing. Tamara F. Lawson, *Can Fingerprints Lie?: Re-Weighing Fingerprint Evidence in Criminal Jury Trials*, 31 AM. J. CRIM. L. 1, 49 (2003).

<sup>172</sup> It should be noted that, while a certain type of error will increase the likelihood of a false result, it can also increase the likelihood of an accurate result. For example, say the defendant is guilty, but a crime scene handwriting sample was not written by the defendant, but by some third party. Further, say the prosecution lab's handwriting analysis falsely finds a match between the crime scene sample and the defendant's handwriting sample. Then, the Type I error actually increases the likelihood of an *accurate* conviction, because even though the evidence is wrong, it points to the correct result. I am assuming that when there is erroneous evidence, the increased likelihood of an erroneous final result is greater than the increased likelihood of an accurate final result. So, the possible societal benefits that could accrue from admitting invalid evidence are not present in my analysis.

only evidence that inculpatates the defendant or exculpates other suspects.<sup>173</sup> If we know which party is offering the expert evidence and the purpose for which it is being offered, we know that in any given proffer of evidence only one type of error, Type I or Type II, is likely.

For example, if the prosecution is offering handwriting analysis evidence that shows a defendant's handwriting sample matches a crime scene note sample, it is impossible that a Type II error occurred in that instance—since there was a positive match between the two samples, there is no chance a false negative occurred. Thus, in evaluating the evidence in that case, we are worried only about the probability of Type I—false positive—error.

The table below maps out this analysis:

	<i>Prosecution Offers Evidence</i>	<i>Defense Offers Evidence</i>
<i>Type I</i> (false positive)	<p><i>Scenario 1</i> Introduces evidence in which samples match, indicating defendant's guilt ↓ Higher likelihood of false conviction</p>	<p><i>Scenario 2</i> Introduces evidence in which samples match, indicating another person's guilt ↓ Higher likelihood of false acquittals</p>
<i>Type II</i> (false negative)	<p><i>Scenario 3</i> Introduces evidence in which samples do not match, indicating another person's innocence ↓ Higher likelihood of false convictions</p>	<p><i>Scenario 4</i> Introduces evidence in which samples do not match, indicating defendant's innocence ↓ Higher likelihood of false acquittals</p>

Of the four scenarios depicted in this table, it seems likely that evidence introduced under Scenario 1—where the prosecution proffers evidence indicating a positive match between the defendant's sample and a police sample—is the most commonly accepted.<sup>174</sup> And so we see that Type I errors—where a defendant's sample is erroneously concluded to match with the police's sample on file—are particularly problematic.

## 2. Option 1: Setting the Error Rate to Type I Error

The matrix reveals two options in defining the error rate for *Daubert* purposes. The first is setting the error rate to the Type I error rate of the scientific methodology.

<sup>173</sup> Whether such evidence is admitted obviously depends on the applicable rules of evidence. The low likelihood of the admission of such evidence does not hinder my analysis—in fact, the analysis hinges on this.

<sup>174</sup> D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99, 104–12 (2000) (contrasting ability of prosecutors' proffered expert testimony to survive virtually all *Daubert* challenges with criminal defendants' expert testimony losing virtually all such challenges).

Looking at the four possible scenarios, Scenario 1—when the prosecution introduces handwriting evidence identifying the defendant—is likely the most prevalent. If the *Daubert* error rate were set to mean Type I errors, Scenario 1 would be better accounted for. And because Scenario 1 is the most prevalent, this definition of the error rate would reflect the relevant error rate in most criminal trial forensic evidence proffers.<sup>175</sup> To extrapolate a general *Daubert* rule from this chart, we want the error rate factor to mean only Type I error.

With regard to the criminal justice goal of consistency, the benefits of this definition are obvious. Defining the error rate as exclusively false positives provides a clear standard for judges to follow. Judges need not make any calculations themselves or ascertain which definition of the error rate is at issue in a particular evidence proffer.

Although relying on this first definition would minimize false convictions, by excluding false negatives, it does not perfectly promote the slightly broader goal of minimizing false results overall. Scenarios 3 and 4 are imperfectly covered by defining the error rate as Type I error since only Type II error is relevant in these cases. The Type I definition fails to discount the possible probative value of expert evidence that falls under Scenarios 3 and 4. In favor of encouraging the second criminal justice goal of clarity and consistency, our first goal of promoting accurate trial outcomes is somewhat compromised.

Defining error rate as only Type I involves another tradeoff: It sets a rule that prioritizes concern for false convictions absolutely over concern for false acquittals. Since we are more concerned with preventing false convictions than with preventing inaccuracy overall (i.e., both false acquittals and convictions), this definition is not fatally problematic. However, one may be slightly concerned with the extreme balance such a rule sets in favor of preventing false convictions over preventing false acquittals.

### 3. *Option 2: A Case-By-Case Definition of the Error Rate*

I propose a second possible definition of the error rate that would strike a different balance in promoting the goals of trial outcome accuracy and consistency. One could devise a more nuanced rule that defines the error rate on a case-by-case basis. This rule would make the judge's considerations of types of error depend on who introduces the evidence and for what purpose: It would define the error rate as the one at issue in that particular proffer of evidence.

For example, imagine you are in Scenario 4: The defense submits handwriting analysis evidence indicating that the defendant's hand-

---

<sup>175</sup> Bernstein, *supra* note 14, at 459.

writing does not match the perpetrator's. Here, the judge must be concerned specifically with a Type II, and not a Type I, error. A more nuanced rule needs to be formulated to account fully for the type of error risked in both Scenarios 1 and 4.

This second "nuanced definition" of the error rate is better than the first at promoting the overall criminal justice goal of trial outcome accuracy by enhancing the probative value of expert evidence. This definition focuses on the type of error at issue in each of all four scenarios. However, it raises a possible problem of impracticability for judges, since it requires judges to ascertain what the relevant error rate definition is for a given proffer of evidence. Thus, this flexible definition of the "error rate" does a worse job than the Type I-only definition of establishing clarity and promoting the goal of trial outcome consistency.

The clarity problem is not necessarily fatal to this second definition, for the rule is not as complicated as it first sounds. It can be reasonably intuitive for a judge to determine which type of error is of concern with any given proffer of forensic evidence: It is the error that increases the likelihood of the intended result of the party proffering the evidence. That is precisely how the error rate rule could be articulated for *Daubert*: The judge should first identify the error rate of the scientific methodology and, if possible, focus specifically on the probability of an error that would increase the likelihood of the result desired by the profferor.

Even with this enunciation of the second definition of "error rate," it is uncertain how much such an articulation would actually promote consistency in criminal trial outcomes. What is certain is that this second "nuanced" definition of the error rate strikes a balance in promoting criminal justice goals that strongly favors an overall concern with trial outcome accuracy over trial outcome consistency. If we remain comfortable with more directly and specifically promoting fairness for defendants and mitigating false convictions, we should adopt the first definition of the error rate and define it as Type I error only.

As a coda, the matrix analysis and the two possible proposed error rate definitions reveal one crucial fact: The ideal error rate definition is never one that means Type I *plus* Type II error. This overstates the possible error for every scenario and does not give an accurate picture of the relevant error rate for any of the four scenarios.

## CONCLUSION

This Note seeks to inspire some hope that doctrinal reforms of the *Daubert* standard could be effective in promoting normative goals of the criminal justice system. My proposed doctrinal tweak should significantly close the resource gap between the prosecution and the defense, directly tackle the problem of forensic evidence's potential for unfair prejudice, and clarify the overly flexible *Daubert* standard. By addressing these problems, this reformulation promotes trial outcome predictability and accuracy and mitigates false convictions.

In contrast to proposed administrative or other systemic solutions, this doctrinal reform is relatively inexpensive. The changes should not significantly increase the costs of *Daubert* hearings. The main costs will be government expenditures to develop specific error rate information for their crime labs through proficiency testing.

This Note clearly proposes only a tweak and not a panacea. However, given the *Daubert* standard's grave shortcomings in criminal trials, even improvements at the margin could make valuable contributions to improving the integrity of our criminal justice system.