
NO ROOM FOR ERROR: CLEAR-EYED JUSTICE IN FORENSIC SCIENCE OVERSIGHT

*Erin Murphy**

The memorandum issued by the Legal Resource Committee to the Organization of Scientific Area Committees (OSAC) convened by the National Institute of Standards and Technology (NIST)¹ is remarkable as much for its existence as for its content. Despite a long history of formal and informal calls to radically restructure forensic science in the United States, the march of reform has been slow and plodding. The emergence of the OSAC, in direct response to the critiques lodged in the 2009 National Academy of Sciences report,² marked one small but meaningful step to prod forensic science away from the strict control of the law enforcement community, and toward a more scientific culture. Indeed, critical space to question forensic orthodoxy has opened simply because the OSAC sits partly within a scientific community (namely NIST) as opposed to wholly within a law enforcement community (such as the Department of Justice or the Federal Bureau of Investigation).³

The Memorandum by the Legal Resource Committee thus represents a rare and welcome development: forensic scientists seeking expert guidance about how to resolve the potential clash between legal and scientific standards applicable to their work. Although asked in connection with the standards for glass comparison, the sophisticated question posed — should the standard for error in forensic testing mirror the aspirational near-zero tolerance for error in criminal adjudication? — is pertinent to a wide array of disciplines. And the answer given by the Committee, approving with caution the use of less demanding error thresholds in forensic testing, likewise applies broadly. But while the Committee's response is almost certainly correct as a matter of law, consideration of the actual implementation of this standard, given the realities of criminal justice practice in the United

* Professor, NYU School of Law.

¹ Memorandum from the Legal Res. Comm. to the Org. of Sci. Area Comms. for Forensic Sci., Nat'l Inst. of Standards & Tech., Question on Hypothesis Testing in ASTM 2926-13 and the Legal Principle that False Convictions Are Worse than False Acquittals (rev. ed. Oct. 7, 2016), *reprinted in* 130 HARV. L. REV. F. 137 (2017) [hereinafter LRC Memo].

² See generally COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009) [hereinafter STRENGTHENING FORENSIC SCIENCE].

³ In many ways, the precursors to the OSAC were the SWGs, or scientific working groups, housed within the FBI. Those groups offered guidance and issued standards for a wide range of forensic disciplines. Although the SWGs remain, the OSAC ended their relative monopoly.

States, reveals the inherent difficulty of the project of forensic reform. Even if factfinders could be made to understand what is at stake in choosing a margin of error, the predominant consumers of this information — judges, defense lawyers, and prosecutors — almost certainly will not, and it will be criminal defendants who suffer as a result.

THE MEMORANDUM, IN THEORY

The question posed to the Legal Resource Committee aimed to clarify whether the law supported findings using scientific standards that grant greater latitude for error than that permitted in the criminal justice system.⁴ In endorsing the use of divergent standards, the Memo distinguished the confidence required to call something a “match” in the statistical sense from the confidence required to assert in a legal forum that two items share a common source (or, ultimately, that the observed similarities support a finding of guilt beyond a reasonable doubt).⁵ In declaring a “liberal [statistical] matching rule . . . legally acceptable,” the Committee pronounced that the “major legal demand” is simply that the witness properly contextualize the strength of the evidence.⁶ Put plainly, a forensic analyst may declare a “match” using a standard that does not require confidence “beyond a reasonable doubt,” so long as the significance of the particular margin of error tolerated is adequately conveyed to legal factfinders.

As a matter of legal reasoning, all of this is probably right. It is an axiom of evidence law that “a brick is not a wall.”⁷ Any single piece of evidence need not prove the ultimate fact conclusively; it can simply be a component of a larger whole that supports a finding of guilt beyond a reasonable doubt. The basic test of evidentiary relevance is simply that a piece of evidence has *any* tendency to make a fact of consequence more or less probable,⁸ a standard that comes with no clear error threshold, but certainly has never been interpreted to impose anything even approximating infallibility. For that reason alone it seems right to conclude that the error bars for scientific findings need not be set so rigidly as to exclude anything other than nearly certain conclusions, or to fix an exceptional threshold for statistical significance. Indeed, on a practical level, such a declaration would trade potentially more valuable and reliable forensic evidence for forms of proof (such as confessions or eyewitness identifications) with much less

⁴ See LRC Memo, *supra* note 1, at 1.

⁵ *Id.* at 6.

⁶ *Id.*

⁷ FED. R. EVID. 401 advisory committee’s note (quoting CHARLES TILFORD MCCORMICK, MCCORMICK ON EVIDENCE § 152, at 317 (Edward W. Cleary ed., 2d ed. 1972)).

⁸ FED. R. EVID. 401.

ascertainable but possibly (and in some cases, demonstrably) higher rates of error.

Nevertheless, when evidence is presented in quantitative terms, or even in sterile language (“match”) imbued with an aura of scientific authority, special safeguards are necessary to ensure accurate transmission and understanding. Our criminal justice model is designed to assess factual reliability through *viva voce* testimony. A lay witness may testify, “That is the robber!” without qualification, because we trust jurors to know that there is an inherent risk of error in such a statement. Moreover, cross-examination or common-sense credibility assessments purportedly work to ferret out any such error. But the notion that factfinders test the credibility and reliability of evidence using their intuitions and observational common sense sits poorly with respect to forensic testimony, which by its nature is recitative rather than spontaneous, and comes with a patina of certainty and infallibility.

Thus, unlike the lay witness, the expert witness should not testify that “the glass shard on the defendant’s coat could have come from the crime scene” without clarifying that, for instance, the match criteria are flimsy enough to render 50% of all clear glass in the world eligible for such a statement of inclusion.⁹ The idea, of course, is that the expert takes on this added burden of clarification because the proposed testimony is scientific in nature, and thus jurors must have the expert’s special knowledge about the strength of match procedures in order to assign the proper weight to the testimony. Indeed, to omit such critical information — perhaps leaving it to cross-examination — is to actively mislead the jury, whereas a lay witness need not spontaneously offer such qualifications. And yet, even with this qualifying statement, there are reasons to doubt whether a factfinder can truly distinguish between “match” testimony by an analyst applying generous margins of error and “match” testimony offered by an analyst applying much stricter standards, as well as whether the forensic analyst has sufficient comprehension to offer such explanations.

The LRC Memo attempts to redress these concerns by underscoring the expert’s obligation to fully explain the limits of certainty. In this respect, the closing sentence of the Memo is key: it notes that an expert’s failure to contextualize the significance of a “match” (including, as I read it, by explaining the breadth of error tolerated) fails to comport with the legal requirement that the expert supply an impartial, reliable opinion on a subject beyond the ken of the ordinary juror.¹⁰ Thus, rather than prescribe a particular error threshold, “[t]he

⁹ Of course, the actual standard suggests a false positive error probability of .003. LRC Memo, *supra* note 1, at 4.

¹⁰ *Id.* at 7.

major legal demand on a scientifically validated and reliable process . . . is that the conclusion be reported with a suitable description of its probative value.”¹¹ If all goes as planned, wider tolerance for error will translate into less powerful evidence in the eyes of the factfinder: that is, “the fact that it is harder for a laboratory whose measurements are imprecise to reject [the null hypothesis (a finding that the items came from different sources)] is balanced by the fact that the laboratory must report a less impressive estimate of the probability of a false inclusion.”¹² With the right explanation, factfinders can make an informed judgment, effectively normalizing the value of any single piece of evidence among methods with varying levels of tolerance for error.

THE MEMORANDUM, IN PRACTICE

But if these underlying theories animate the Committee’s approach, then a word about their limitations may be in order. The point here is not to reread enduring debates about “hard” versus “soft” forms of proof.¹³ Rather, it is to suggest three reasons why the aspirations of the LRC Memo’s announced standard — while the right one to make “on paper” — may fail to be realized in the sociocultural environment of forensic practice, to detrimental consequence. Specifically, I argue that: 1) factfinders will be largely incapable of assigning meaningful weight to evidence according to the subtle methodological differences described; 2) given the realities of our system, it is not really factfinders who will undertake this difficult task, but rather defense lawyers and prosecutors, yet the culture of forensic evidence in the criminal justice system does not currently encourage the clear transmission of this information; and 3) criminal defendants will in large part be the ones who suffer as a result of the application of liberal match rules.

First, the LRC Memo, in deciding whether wider margins of error than those that would be permitted in “reasonable doubt” standards are tolerable in statistical calculations, elects an explanatory approach over one that would set a specific threshold of statistical significance before admission of a match statement. But the explanatory route works only if jurors can properly hear and weigh the explanation given. A long string of studies suggests that jurors dramatically alter their estimation of the import of the evidence based on simple changes in the way in which experts phrase their testimony, with qualitative

¹¹ *Id.* at 6.

¹² *Id.*

¹³ See, e.g., Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1365 (1971); Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245 (2016).

statements of evidentiary value tending to carry the most weight.¹⁴ More to the point, studies of how jurors incorporate error rate information generally indicate that they do so poorly, even when efforts are made to help them assimilate a probability and its associated confidence level.¹⁵ Thus, if an expert may testify that the suspect's sample "matches" or "could have come from the source," then research suggests that the jury will internalize these ideas, either minimizing or dismissing outright any added qualifying statements as to the statistical uncertainty underlying those assertions. In short, what matters most to jurors is the "match" statement — whether qualitative or quantitative; jurors do not seem to mire themselves in the details of the scientific standards applied to determine what constitutes a match.

Of course, a logical response to this concern would be to cite that great engine of truth: cross-examination by defense counsel. The trouble with that — setting aside the aforementioned studies that indicate that even concerted efforts to educate fail — is that the actual practice of criminal defense does not always live up to its ideals either. Indigent defense lawyers are notoriously overworked and underpaid, and many lack basic competencies, much less sophisticated scientific expertise.¹⁶ Trials are rare; the vast majority of cases are resolved in plea

¹⁴ See, e.g., Jonathan J. Koehler, Audrey Chia & Samuel Lindsey, *The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?*, 35 JURIMETRICS J. 201, 212–14 (1995) (finding that jurors could not make sense of an error rate, and were instead dazzled by a seemingly rare probability); Dawn McQuiston-Surrett & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 LAW & HUM. BEHAV. 436 (2009); Nicholas Scurich & Richard S. John, *Mock Jurors' Use of Error Rates in DNA Database Trawls*, 37 LAW & HUM. BEHAV. 424, 429 (2013) ("The current study extends this line of research to a new domain, namely, database trawl cases, and finds, consistent with the majority of previous studies, that jurors underappreciate the relative effect of error rates on probative value."); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?*, 52 LAW & CONTEMP. PROBS. 9, 38 (1989) (concluding that "under some circumstances jurors may seriously overestimate the value of statistical evidence" due to "their tendency to give equal weight to statistical evidence that varies widely in its probative value"). But see Dale A. Nance & Scott B. Morris, *Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability*, 34 J. LEGAL STUD. 395, 410 (2005) (finding that error rate testimony yielded no appreciable difference in appraisal of evidence).

¹⁵ These studies tend to focus on laboratory error rate, as opposed to methodological errors. Jason Schklar & Shari Seidman Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 LAW & HUM. BEHAV. 159, 159 (1999) (reporting findings of study); see also *id.* at 163 (describing vividness hypothesis, which holds that jurors become mesmerized by seemingly powerful statistics, which swallow less memorable error rate information, and the averaging strategy hypothesis, which holds that jurors simply incorporate each piece of information erroneously).

¹⁶ THOMAS GIOVANNI & ROOPAL PATEL, BRENNAN CTR. FOR JUSTICE, *GIDEON AT 50: THREE REFORMS TO REVIVE THE RIGHT TO COUNSEL* 4–5 (2013), http://www.brennancenter.org/sites/default/files/publications/Gideon_Report_040913.pdf [<https://perma.cc/KYV8-C23S>] (describing working conditions and lack of training).

proceedings. Forensic reports often do not even make it to defense counsel before a plea is negotiated; live testimony is rarer still.

This leads us to the second point, which is that the image of evidence being tested through an adversarial crucible is pure fantasy in the overwhelming number of cases. Instead, the evidence is most often accepted at face value, based on the documents given to the defense attorney by the prosecutor, who typically serves as the analysts' intermediary. If that is the reality of forensic evidence — not the lofty ideals of the criminal trial with informed defense counsel conducting a vigorous and engaging cross, but the rushed plea negotiation after a quick flip through the discovery — then the content of an analyst's written report is, more than anything, what will determine whether adequate distinction is made between strong and weak evidence.

In this respect, the LRC Memo's demand for a "suitable description" of the evidence's probative value feels tepid. The current culture of forensic science is still heavily entwined with law enforcement interests, despite gradual progress toward greater independence and professionalism that largely began in the wake of the 2009 National Academy of Science's report.¹⁷ Virtually no standards for report writing exist; local culture tends to dominate. If a "liberal matching rule" is to comport with the ideals of justice, then the "suitability" of the description of its liberality is critical. Without analysts trained and acculturated to revealing the degree to which their methodologies tolerate error, and to writing reports that spell out the implications of those choices in plain English, the choice of a particular threshold of statistical significance will have profound effects on the mechanisms of justice.¹⁸ Rather than nest within the Memo a single line encouraging greater clarity, the Committee could have explicitly conditioned tolerance for relaxed error margins on reports that contain express, clear statements as to the meaning and impact of different margins.

Third, and finally, refusing to set an absolute (and demanding) threshold for statistical significance will, as regards forensic evidence, almost always hurt the criminal defendant. The Memo points out that defendants might themselves marshal forensic proof in their defense, or that such evidence might arise in civil cases.¹⁹ But instances of this kind will be wholly dwarfed by the standard use of forensic proof: to inculcate a criminal defendant. Moreover, it is wrong to suggest that

¹⁷ See generally STRENGTHENING FORENSIC SCIENCE, *supra* note 2.

¹⁸ If anything, this is the strongest argument in favor of admitting evidence only when produced by conservative thresholds. In an ideal world, factfinders, lawyers, and defendants could receive the information about confidence levels and properly incorporate it into their decisionmaking, but in reality that may be too much to ask. Thus, it is perhaps defensible to simply foreclose admission of "match" evidence that does not meet the highest levels of certainty.

¹⁹ LRC Memo, *supra* note 1, at 5.

exposing the criminal defendant to injustice is a necessary cost of endorsing a universal standard.

After all, nothing prohibits setting a strict standard for forensic evidence offered in a criminal case, while permitting more liberal standards in the civil context.²⁰ Pattern match testimony in a car accident case may defensibly be more readily admissible than that in a homicide. To the extent that application of less stringent match criteria seems odd, it is easily explicable as consistent with the lesser burden of proof applied in civil cases more generally. Regardless, even a unitary standard set high in order to respect the seriousness of criminal matters would not exclude a tremendous amount of evidence in civil cases: most forensic methods have little crossover to civil contexts — the chief exceptions are handwriting analysis and fire science.²¹

In addition, to the extent that the concern is that a criminal defendant might not be able to use exculpatory evidence based on such high thresholds, the Constitution might intercede to allow otherwise inadmissible proof.²² Of course, in actuality there already exist significant impediments to a defendant introducing evidence that exculpates by means of inculpating another person. Courts tend to apply a heightened standard of relevance before admitting evidence that purports to point toward a third-party perpetrator. One commenter has observed that such impediments are so severe that, “[i]n most jurisdictions in the real world when a defense attorney attempts to assay a SODDI [“some other dude did it”] defense, he or she encounters substantial hurdles.”²³ That is the primary reason why exculpatory forensic evidence is unlikely to be used: because its main useful purpose would be to point to another perpetrator, but ordinary evidence rules largely preclude admission of such evidence.

Lastly, to the extent that the defendant proffers such evidence as exculpatory because it *fails* to match demanding significance standards, this chiefly poses a risk of wrongly exculpating a guilty person (that is, a Type II/false negative error), not wrongly convicting an in-

²⁰ The existing Federal Rules of Evidence occasionally distinguish between admissibility and inadmissibility on the basis of whether the evidence is offered in a criminal case. See, e.g., FED. R. EVID. 404, 609.

²¹ See, e.g., Erin Murphy, *Neuroscience and the Civil/Criminal Daubert Divide*, 85 *FORDHAM L. REV.* 619, 621–24 (2016).

²² Cf. *Rock v. Arkansas*, 483 U.S. 44, 49–52 (1987). For instance, strict match criteria could result in a “no match” finding, but a defendant might argue that application of a relaxed standard would produce an exculpatory “match.” In such a case, a court could simply judge the more liberal criteria sufficiently reliable and admit the evidence in keeping with the defendant’s constitutional rights.

²³ Edward J. Imwinkelried, *Evidence of a Third Party’s Guilt of the Crime that the Accused is Charged with: The Constitutionalization of the SODDI (Some Other Dude Did It) Defense 2.0*, 47 *LOY. U. CHI. L.J.* 91, 93 (2015); see also *id.* at 97–98 (describing majority position of applying heightened standards for admissibility of such evidence).

nocent (that is, a Type I/false positive error). Such errors may be cause for concern, but they generally do not run afoul of the overarching principle that legal standards in criminal justice ought to be set such that they tolerate virtually no error.

THE MEMORANDUM, REVISED

Imagine that the Legal Resource Committee took a bolder, more holistic stance in responding to the question asked. Set aside the option of simply concluding that, as a result of structural infirmities in the criminal justice system, the thresholds for error between science and law must align. Although defensible, such an approach would arguably cause serious injustice by excluding valuable evidence (including whole categories of proof that could never satisfy such high standards), not to mention meet with vigorous resistance from government actors.

Consider the ways in which the LRC Memo could have refused to answer the question in isolation, instead locating it directly within the realities of criminal justice. Such a response would have included more texture about the everyday use of forensic evidence by criminal justice actors,²⁴ and more aggressively established the link now hazily drawn between legal tolerance of lower error thresholds and legal dictates to convey the implications of those thresholds. In this view, the Memo could have conditioned its reply on prescribed reporting standards, such as requiring detailed, plain-spoken language explaining the significance of chosen error rates; additional precautions to minimize the possibility for cognitive biases; and demands that analysts who employ thresholds short of reasonable doubt adhere to the discovery and reporting recommendations laid out by the National Commission on Forensic Science.²⁵ By linking these systemic remedies with the question asked, the Memo could have infused as much *legal practice* as *law* into the culture of forensic science, which would be a welcome development indeed.

In sum, the LRC Memo is indisputably correct that the law does not dictate that the forensic match testimony must be withheld unless

²⁴ Such texture is readily available in a variety of scholarly accounts. See, e.g., Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence*, 95 CALIF. L. REV. 721 (2007); NAT'L RESEARCH COUNCIL, *supra* note 2.

²⁵ See Memorandum from the Nat'l Comm'n on Forensic Sci., Nat'l Inst. of Standards & Tech., to the Attorney General, Pretrial Discovery (June 21, 2016), <https://www.justice.gov/ncfs/file/880241/download> [<https://perma.cc/PB9L-U2FU>]; Memorandum from the Nat'l Comm'n on Forensic Sci., Nat'l Inst. of Standards & Tech., to the Attorney General, Use of the Term "Reasonable Scientific Certainty" (Mar. 22, 2016), <https://www.justice.gov/ncfs/file/839726/download> [<https://perma.cc/J3EB-P3ES>]; Memorandum from the Nat'l Comm'n on Forensic Sci., Nat'l Inst. of Standards & Tech., to the forensic science community, Inconsistent Terminology (Apr. 30, 2015), <https://www.justice.gov/ncfs/file/477841/download> [<https://perma.cc/6XPH-QDCV>].

the methodology applied has a level of confidence akin to the high threshold imposed for the ultimate finding of guilt. And the Memo is right to take pains to underscore that part of the ethical and legal (in the sense of admissibility standards) obligation of an expert is to make clear to the jury what the precise value of the evidence is, including its limitations or possible shortcomings. But the Memo unfortunately glosses over the strong probability that 1) forensic analysts fail to meaningfully convey information about the strength or weakness of their standards in the documents they provide in discovery; 2) prosecutors and defense attorneys, in the crush of caseloads that barely allow time for assessment of evidence, fail to recognize its value even when they do; and 3) this will hurt criminal defendants, who are more likely to accept pleas that they should not or receive sentences that are unjust as a result of misunderstood forensic evidence. Without addressing those three realities of criminal process, the underlying problem of misuse of forensic evidence will remain. And so the cycle continues.