Journal of Law and Policy

Volume 21
Issue 2
SYMPOSIUM:
Authorship Attribution Workshop

Article 11

2013

Linguistic Confusion in Court: Evidence From the Forensic Sciences

Jonathan J. Koehler, Ph.D.

LINGUISTIC CONFUSION IN COURT: EVIDENCE FROM THE FORENSIC SCIENCES

*Jonathan J. Koehler**

INTRODUCTION

When fingerprint evidence was approved for admission in U.S. courts in 1911,[1] the approving court noted that "[e]xpert evidence is admissible when the witnesses offered as experts have peculiar knowledge or experience not common to the world, which renders their opinions, founded on such knowledge or experience, an aid to the court or jury in determining the questions at issue."[2] In other words, expert testimony is appropriate when a qualified witness has something to say that helps a fact finder in the instant case. One hundred years later, the sentiment expressed in *Jennings* appears in the Federal Rule of Evidence pertaining to the admissibility of expert testimony.[3] But how can we know when expert testimony is helpful rather than unhelpful or even harmful?

I submit that the *specific language* used in court by experts can be the difference between testimony that is truly helpful and testimony that is confusing or unhelpful. This idea is particularly germane to scientific testimony in cases where the triers of fact have a limited understanding of the principles and methods from which the testimony derives its strength.

---

* Beatrice Kuhn Professor of Law, Northwestern University School of Law.

[1] People v. Jennings, 96 N.E. 1077, 1081–82 (Ill. 1911).

[2] *Id.* at 1083.

[3] FED. R. EVID. 702(a) (stating that expert testimony is admissible when it "will help the trier of fact to understand the evidence or to determine a fact in issue").

This simple point, which has obvious implications for the law, should also be of interest to the forensic linguistics community. Forensic linguists are sometimes called to testify about document content analysis, speaker identification, and authorship, among other things. How, for example, should forensic linguists testify about their analyses of who wrote a particular text message? At present, forensic scientists in general, and forensic linguists in particular, take one of two very different approaches. One approach is testimony that culminates in the expert's *subjective source opinion*. For example, those who practice "forensic stylistics" commonly offer their opinions about who wrote (or who did not write) a document after taking account of such stylistic characteristics as document format, spelling, capitalization, abbreviations, punctuation, word choice, and syntax.[4] Among the more established forensic sciences, fingerprint analysis offers a similarly subjective conclusion.[5] The second approach culminates in a *quantitative statement about the degree of observed correspondence* between an unknown target and a known reference. In forensic linguistics, this approach is favored by computational linguists because it fits well with the field's tendency to identify statistical models for language use. However, there are few, if any, databases from which to generate quantitative statements. DNA analysis, which does rely on large databases to generate probability assessments, offers similarly data-driven probabilistic conclusions. For example, a DNA analyst will commonly report the frequency with which a matching DNA profile exists in a reference population (i.e., the "random match probability").[6]

Regardless of which approach is used at trial, there is a significant risk that expert testimony on scientific and technical matters will confuse or mislead triers of fact.[7] This risk is particularly important whenever material is unfamiliar or

---

[4] ANDRE E. MOENSSENS ET AL., SCIENTIFIC EVIDENCE IN CIVIL AND CRIMINAL CASES 252 (6th ed. 2013).

[5] NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 139 (2009) ("[F]riction ridge analysis relies on subjective judgments by the examiner.").

[6] MOENSSENS ET AL., *supra* note 4, at 862.

[7] FED. R. EVID. 403.

complex. In these cases, the specific language used by legal actors may be the difference between testimony that is truly helpful to the trier of fact and testimony that is misleading and unhelpful.[8] In this paper, I address issues related to how scientific and technical information should and should not be communicated in court. Because confusion in the DNA and fingerprint areas has been documented and is relatively common, my comments focus on linguistic problems in DNA and fingerprint expert testimony in hopes that forensic linguists can avoid the testimonial traps and errors that plague these forensic scientists.

Section I of this paper examines DNA match statistics and describes the confusion that legal actors experience when dealing with conditional probabilities. Section II examines statistical inverse errors in the 2010 U.S. Supreme Court case *McDaniel v. Brown*.[9] Section III examines a seductive, but faulty, statistical assumption that commonly arises in paternity cases. Section IV examines the role of error rates in forensic sciences and concludes that identifying those error rates is particularly important in fields that rely on highly discriminating statistical techniques. Section V offers an illustration of the crucial role ill-defined language can play in a legal proceeding. Standard and precise terms are recommended. The paper concludes with a section identifying implications for the forensic linguistics and authorship attribution communities.

I. DNA MATCH STATISTICS

When an expert testifies about DNA evidence found at a crime scene, the punch line is usually statistical. Specifically, after the expert testifies to a "match" (or inclusion) between the DNA profile of an evidentiary sample and the DNA profile of a particular individual, the strength of that match is often described by the random match probability ("RMP"). The RMP

---

[8] The Federal Rules of Evidence require that expert testimony must "help the trier of fact to understand the evidence or to determine a fact in issue." FED. R. EVID. 702(a).

[9] McDaniel v. Brown, 130 S. Ct. 665 (2010).

is a statistic that describes the frequency of a DNA profile in a population.[10] Other things being equal, smaller RMPs (such as one in one billion) indicate a stronger DNA match than larger RMPs (such as one in one hundred) because the chance that the match is purely coincidental in the former instance is much less likely.[11]

---

[10] MOENSSENS ET AL., *supra* note 4, at 863 ("[T]he 'random match probability' (RMP) is the probability that a randomly selected, unrelated individual in the relevant population would have a particular DNA profile.").

[11] Importantly, there are at least two circumstances in which the RMP provides a misleading indicator of the strength of a DNA match. The first circumstance is when the potential source population includes close relatives of the putative source. The chance that a putative source will share a DNA profile with a close relative is usually much larger than the RMP, and therefore the chance of a coincidental match with the crime scene sample is larger as well. *See* NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., THE EVALUATION OF FORENSIC DNA EVIDENCE 123 (1996). The second circumstance in which the RMP provides a misleading indicator of the strength of a DNA match is when the risk of laboratory error is substantially larger than the RMP. COLIN AITKEN & FRANCO TARONI, STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENCES 425 (2004) ("If the probability of an error . . . is much greater than the probability of matching profiles . . . then the latter probability is effectively irrelevant to the weight of the evidence."); DAVID J. BALDING, WEIGHT-OF-EVIDENCE FOR FORENSIC DNA PROFILES 35 (2005) ("If the false-match probability (ii) is judged to be much larger than the chance-match probability (i), then the latter probability is effectively irrelevant to evidential weight . . . . [I]t is not the absolute but the *relative* magnitude of the false-match to the chance-match probabilities that determines whether the former can be safely neglected."); Jonathan J. Koehler et al., *The Random Match Probability (RMP) in DNA Evidence: Irrelevant and Prejudicial?*, 35 JURIMETRICS J. 201 (1995) ("RMPs contribute little to an assessment of the diagnostic significance of a reported DNA match beyond that given by the false positive laboratory error rate when RMPs are several orders of magnitude smaller than this error rate."); Richard Lempert, *After the DNA Wars: Skirmishing with NRC II*, 37 JURIMETRICS J. 439, 447 (1997) ("the probative value of a DNA match is always limited by the chance of false positive error"); William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 J. FORENSIC SCI. 1, 1 (2003) ("[H]aving accurate estimates [of] the false positive probabilities can be crucial for assessing the value of DNA evidence."). Laboratory error includes all types of error that might result in a reported match on a person who is not, in fact, the source of the evidentiary item.

The strength of a DNA match may also be given by a likelihood ratio ("LR").[12] A LR is a ratio of conditional probabilities that examines the probability of observing evidence under two competing hypotheses.[13] The LR technique allows experts to determine how much more (or less) the evidence favors one hypothesis over the other. Ignoring for the time being the twin issues of close relatives and laboratory error identified in footnote 10, the LR is approximately the inverse of the RMP (i.e., 1/RMP).[14] The numerator is approximately 1 (or 100%) because if the putative source is, in fact, the actual source of the evidentiary item, then he or she will share a common DNA profile with the evidentiary item.[15] Similarly, if the putative source is not, in fact, the actual source, then he or she will not share a DNA profile with the true source unless he or she, by sheer coincidence, has the same DNA profile.[16] The RMP captures the chance of a coincidental match and is commonly

---

[12] AITKEN & TARONI, *supra* note 11, at 153–55.

[13] NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *supra* note 11, at 31. ("The LR is the ratio of the probability of a match if the DNA in the evidence sample and that from the suspect came from the same person to the probability of a match if they came from different persons.").

[14] *Id.* ("Since the probability of a match when the samples came from the same person is one (unless there has been a mistake), the likelihood ratio is simply the reciprocal of the match probability.").

[15] This assumes, of course, that a person's DNA profile remains constant across time. In rare cases, an individual's DNA may change. *See, e.g.*, Cai Wenjun, *Rare Mutation Solves Murder*, SHANGHAI DAILY (Nov. 12, 2012), http://www.shanghaidaily.com/nsp/National/2012/11/12/Rare%2Bmutation%2Bsolves%2Bmurder/ (discussing a rare mutation that aided police in identifying a suspect from a pair of identical twins).

[16] The chance of a coincidental match depends, in large part, on how many loci are examined. Today, thirteen loci are most commonly used, and the resultant random match probabilities are on the order of one in billions, trillions, and quadrillions. *See* NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *supra* note 5, at 3-12; NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 142 (3d ed. 2011), *available at* http://www.au.af.mil/au/awc/awcgate/fjc/manual_sci_evidence.pdf; WILLIAM C. THOMPSON, THE POTENTIAL FOR ERROR IN FORENSIC DNA TESTING (AND HOW THAT COMPLICATES THE USE OF DNA DATABASES FOR CRIMINAL IDENTIFICATION) 6–7 (2008), *available at* http://www.councilforresponsiblegenetics.org/pageDocuments/H4T5EOYUZI.pdf.

inserted into the denominator of the LR.[17] Thus, when the RMP is 1 in 3,000,000, the corresponding LR is often reported as 3,000,000:1. This means that the matching DNA profile is 3,000,000 times more likely under the hypothesis that the defendant is the source of the evidentiary item than under the hypothesis that the defendant is not the source.

What this does not mean, however, is that the defendant is 3,000,000 times more likely to be the source of the evidentiary item than not to be the source. Most people, experts included, would be hard-pressed to explain why this is so. But a careful review of the relevant conditional probabilities provides insight. The LR describes P(Evidence | Source) / P(Evidence | Not Source). However, the statement "the defendant is 3,000,000 times more likely to be the source of the evidentiary item than not to be the source," describes the posterior odds ratio P(Source | Evidence) / P(Not Source | Evidence). The posterior odds ratio is the inverse of the LR. Those who confuse the LR with the posterior are committing a transposition error or "inverse fallacy."[18] This error is no mere technicality. Just as we may not assume that the probability that Jack will eat a hot dog given that he is at the ball game (very high probability) is the same as the probability that Jack is at a ball game given that he is eating a hot dog (much lower probability), we may not assume that P(Source | Evidence) = P(Evidence | Source) or that P(Not Source | Evidence) = (Evidence | Not Source).

Nonetheless, people often commit inverse errors when dealing with conditional probabilities.[19] People also confuse

---

[17] NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., s*upra* note 11, at 31.

[18] D.H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?*, 154 J. ROYAL STAT. SOC'Y SERIES A 75, 77–78 (1991).

[19] Ward Cascells et al., *Interpretation by Physicians of Clinical Laboratory Results*, 299 NEW ENG. J. MED. 999, 1001 (1978) (showing 45% inverse errors among Harvard physicians); Leda Cosmides & John Tooby, *Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty*, 58 COGNITION 1, 25 (1996) (showing 56% inverse errors among Stanford students); Kaye & Koehler, *supra* note 18, at 77 (reviewing inversion fallacy data in pre-DNA mock juror studies conducted in the 1980s); Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies,*

conditional probabilities with joint probabilities,[20] and are less likely to engage in sound probabilistic reasoning when using conditional probabilities than when those probabilities are converted into frequency form.[21] These problems may have significant consequences for legal cases that involve scientific and statistical testimony. Jurors who make these mistakes may believe that the RMP identifies the probability that the defendant is innocent. This belief is known as the "prosecutor's fallacy."[22] There is evidence that experts, attorneys, and other legal actors fall prey to this fallacy in actual cases.[23] Similarly, legal actors fall prey to the source probability error,[24] which involves equating the RMP with the probability that the putative source is not the source of the evidentiary item in question. This latter error is so tempting that the RMP is *routinely* described in court

---

Likelihood Ratios and Error Rates, 67 COLO. L. REV. 859, 877–78 (1996) (noting that people treat LRs much as they treat posterior odds ratios); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence*, 52 LAW & CONTEMP. PROBS. 9 (1989); Christopher R. Wolfe, *Information Seeking on Bayesian Conditional Probability Problems: A Fuzzy-Trace Theory Account*, 8 J. BEHAV. DECISION MAKING 85, 97 (1995) (noting that 77% of college students verbally confused LRs with posterior odds ratios).

[20] Stephen E. Edgell et al., *Base Rates, Experience and the Big Picture*, 19 BEHAV. & BRAIN SCI. 21, 21 (1996); Gerd Gigerenzer & Ulrich Hoffrage, *How to Improve Bayesian Reasoning Without Instruction: Frequency Formats*, 102 PSYCHOL. REV. 684, 694 (1995).

[21] Cosmides & Tooby, *supra* note 19, at 25 (comparing errors among Stanford students and finding a 56% rate for inverse errors but only 5% rate when frequencies used); William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167, 172–76 (1987) (noting that 22% committed inverse fallacy on blood matching evidence in the context of a hypothetical robbery case when the evidence was presented in P(E | -G) form, whereas a frequency presentation of the blood evidence produced inverse fallacies only 4% of time).

[22] Thompson & Schumann, *supra* note 21, at 171.

[23] McDaniel v. Brown, 130 S. Ct. 665, 672–73 (2010); Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence*, 34 JURIMETRICS J. 21, 32 (1993).

[24] *McDaniel*, 130 S. Ct. at 673; AITKEN & TARONI, *supra* note 11, at 81–82; Koehler et al., *supra* note 11, at 212.

opinions as the chance that someone other than the defendant is the source of the genetic evidence.[25]

II. STATISTICAL INVERSE ERRORS: *MCDANIEL V. BROWN*

Probability errors took center stage, at least in defense filings and an amicus brief,[26] in the U.S. Supreme Court case *McDaniel v. Brown*.[27] In *McDaniel*, Troy Brown was tried and convicted of a rape in Nevada largely based on DNA evidence. Renee Romero, a criminalist for the county, discovered semen on the victim's underwear that matched Brown's DNA profile. On direct exam, Romero estimated the frequency of the DNA profile to be "one in 3 million."[28] When the prosecutor asked "[s]o that means that only one in 3 million people will share the same genetic code?" Romero correctly answered in the affirmative.[29]

The Supreme Court described Romero's testimony on this matter as follows: "The State's expert, Renee Romero, tested the [blood stain] and determined that the DNA matched Troy's and that the probability another person from the general population

---

[25] State v. Reaves, No. COA10–1246, 716 S.E.2d 441, at *3 (N.C. Ct. App. Oct. 4, 2011) (unpublished table decision) ("The lowest probability that someone other than Defendant in the North Carolina African American population contributed the DNA discovered on Ms. Curtis' steering wheel was one in 147,000."); State v. Timm, No. 13–11–23, 2012 WL 367589, at *2 (Ohio Ct. App. Feb. 6, 2012) ("Mr. Weiss testified that the statistical probability that someone other than Timm could be the source of the DNA in the sperm fraction extracted from the shorts was less than one in more than 6.5 billion."); Murga v. State, No. 05–10–01237, 2012 WL 807081, at *2 (Tex. Ct. App. Mar. 13, 2012) ("The third analysis showed a one in 11.1 billion possibility that someone other than appellant had a DNA profile that matched appellant's.").

[26] Brief for 20 Scholars of Forensic Evidence as Amici Curiae Supporting Respondents, *McDaniel*, 130 S. Ct. 665 (No. 08-559), *reprinted in* Erin Murphy & William C. Thompson, *Common Errors and Fallacies in Forensic DNA Statistics: An Amicus Brief in* McDaniel v. Brown, 46 CRIM. L. BULL. 5 (2010).

[27] *McDaniel*, 130 S. Ct. at 671.

[28] Jury Trial Transcript Day 3, September 29, 1994, State v. Brown, No. 5833 (Nev. Dist. Ct. 1994), *reprinted in* 2 Joint Appendix at 330, 437, *McDaniel*, 130 S. Ct. 665 (No. 08-559) [hereinafter McDaniel Transcript].

[29] *Id.* at 438.

would share the same DNA (the 'random match probability') was only 1 in 3,000,000."[30] The Court's characterization of Ms. Romero's testimony is ambiguous. When the Court says, "the probability another person from the general population would share the same DNA . . . was only 1 in 3,000,000,"[31] it might mean (a) the chance that *any* person in the general population (The U.S.? The world?) would share the DNA profile in question is 1 in 3,000,000, or (b) the *frequency* with which people in the general population share the DNA profile in question is 1 in 3,000,000. The latter interpretation is the correct interpretation of what Ms. Romero actually said,[32] but one cannot know this from the Court's imprecise language. The ambiguity in the Court's restatement here is ironic given that a central issue raised by the defendant in the appeal was the use of imprecise language concerning the DNA evidence at trial and its impact on those who heard it.

A. Source Probability Error

In all likelihood, the prosecutor in *McDaniel* wanted the expert to state the RMP in a more dramatic fashion. That is, he probably wanted Romero to describe it as a source probability. Of course, the RMP cannot be converted to a source probability. But the prosecutor nonetheless attempted to extract one from Romero. This following exchange between the prosecutor and Romero ensued:

> Q: Now, as far as a—for my benefit, we're looking at a one in 3 million statistic. Is there another way to show that statistic? In other words, what—let's say 100 percent—what is the likelihood that the DNA found in the panties is the same as the DNA found in the defendant's blood?

---

[30] *McDaniel*, 130 S. Ct. at 668.

[31] *Id.*

[32] We know that the latter interpretation is the right interpretation of what Romero said because she agreed with the prosecutor that the one in three million statistic meant that "only one in 3 million people will share the same genetic code." McDaniel Transcript, *supra* note 28, at 438.

A: Paternity testing uses percentages.

Q: Okay.

A: Not the way forensics likes to look at it. We prefer the one in 3 million.

Q: I understand that, but for just another way to look at it, what would that percentage be?

A: It would be 99.99967 [*sic*][33] percent.[34]

When the prosecutor asks "what is the likelihood that the DNA found in the panties is the same as the DNA found in the defendant's blood?" he appears to be asking for a source probability. Though hard to interpret, he seems to want Ms. Romero to identify the probability that the DNA in the panties and the DNA in the defendant's blood share a common source. In other words, he seems to be asking Ms. Romero to identify P(Defendant is the Source of the Recovered DNA | The Recovered DNA Matches the Defendant). Ms. Romero's initial answer—"Paternity testing uses percentages"—is not responsive. Ms. Romero's second answer—"we prefer the one in 3 million," is odd for several reasons. First, "one in 3 million" is a frequency, not a percentage. This contradicts her immediately preceding statement about using "percentages" in paternity testing. Second, it is a simple matter to translate a frequency into a probability. Here, for example, a frequency of one in 3,000,000 is mathematically identical to .000033%.

When the prosecutor presses Ms. Romero further by asking for "another way to look at it," the "it" transforms from an RMP of 1 in 3,000,000 (or .000033%) to "1 – RMP" or 99.999967%.[35] Obviously .000033% is not the same as 99.999967%. Therefore, it is apparent that Ms. Romero was not referring to the RMP at all when she offered the 99.999967% figure.

Having succeeded in extracting the 99.999967% source probability estimate from his expert, the prosecutor next attempts

---

[33] According to the *McDaniel* transcript, Ms. Romero said "99.99967 percent." *Id.* at 458. However, she presumably meant (or said) "99.999967 percent" which is the outcome of 100 percent minus .000033 percent.

[34] *Id.*

[35] *Id.*

to get Romero to restate the one in three million RMP as the probability that the defendant is not the source of the DNA, i.e., P(Not Source | Match).

> Q: So, the—would it be fair to say, based on that that the chances that the DNA found in the panties—the semen in the panties—and the blood sample, the likelihood that it is not Troy Brown would be .000033?
>
> A: I'd prefer to refer to it as the one in 3 million.
>
> Q: All right. But from a mathematical standpoint, would that be inaccurate?
>
> A: Repeat the question, please.
>
> Q: Would it be fair, then, to say that with that mathematical calculation there, that the likelihood that the DNA extracted from the semen in the panties and the DNA extracted from the blood that the likelihood that it's not Troy Brown, that it's not a match is .000033?
>
> Mr. Lockie [Defense Counsel]: Your Honor, I'm going to object on relevance. The witness is testifying that it's not scientifically valid in her opinion. So it's not relevant.
>
> The Court: Well, I don't know that—
>
> [Mr. Smith (Prosecutor)][36]: That's just a subtraction problem.
>
> The Court: Let's go back. I don't think that's what she said. I don't think that's what she said. Let's go back a step and find out. I don't think that's what she said.
>
> By Mr. Smith [Prosecutor] (continuing)
>
> Q: Now, I understand that—and what I'm trying to do is make this into a percentage where I can understand it. And so I recognize that as far as your testing, you would prefer to have it as a one in 3 million, but just as another

---

[36] The trial transcript indicates that Mr. Lockie (defense counsel) makes this comment, but it seems unlikely that he would contradict his own objection by stating that this is "just a subtraction problem." *Id.* at 460–61. The statement was probably made by the prosecutor in response to defense counsel's objection as indicated in the text above.

> way of looking at it, would it be inaccurate to state it that way?
>
> A: It's not inaccurate, no.
>
> Q: All right. Then in response to my question, would the likelihood that the semen from the DNA found in the panties and the blood from Troy Brown, that it's not the same, would it be—the chances that they are not a match would be .000033?
>
> A: Yes. That's the way the math comes out.
>
> Q: All right.
>
> THE COURT: Let's make sure. It's the same thing—it's the same math just expressed differently. Is that correct?
>
> THE WITNESS: Yes. Exactly, your Honor.
>
> THE COURT: Thank you.[37]

As before, Romero initially resists the prosecution's efforts to turn an RMP into a source probability by stating a preference for expressing the DNA statistic as a frequency rather than as a probability. But Romero's resistance misses the mark. As noted above, it makes no mathematical difference whether a frequency statistic is expressed as a frequency or as its equivalent probability (decimal) value.[38] One in three million may be described as .00000033 or as its percentage equivalent, .000033%.

What Romero *should have* resisted was the prosecutor's attempt to convert the .000033% RMP statistic[39] into a posterior

---

[37] *Id.* at 460–62.

[38] Of course, although frequencies and their corresponding probabilities are mathematically equivalent, people may respond differently to the form of the presentation. Indeed, there is evidence that people respond differently to frequencies and their mathematically equivalent probabilities in the context of DNA statistics. *See generally* Jonathan J. Koehler & Laura Macchi, *Thinking About Low-Probability Events: An Exemplar-Cuing Theory*, 15 PSYCHOL. SCI. 540 (2004) (finding that people were less persuaded by low probability DNA evidence when it was presented in an exemplar-conducive way than when it was not).

[39] The prosecutor omits the "percent" on the .000033% RMP statistic. Although surely unintentional, this omission inflates the RMP from one in 3,000,000 to one in 30,000. McDaniel Transcript, *supra* note 28, at 460–62.

probability that the matchee is not the source of the evidence. The conversion of an RMP into a posterior probability is not simply "another way of looking at it,"[40] as the prosecutor suggested. It is a fallacious maneuver, albeit one that the prosecutor may not have realized was fallacious. Indeed, if one were to assign blame for the statistical confusion on this point, it must fall squarely on the shoulders of the expert witness, Romero. When the prosecutor committed the inverse fallacy and then directly asked Romero, "[W]ould it be inaccurate to state it that way?"[41] Ms. Romero erroneously replied, "It's not inaccurate, no."[42] She affirmed this error repeatedly in this exchange both with the prosecutor and then with the trial judge.[43]

We should expect more from forensic science experts who offer statistical testimony. They must know what the inverse fallacy is, they must correct the error when it is made by judges or attorneys, and they certainly must not promote it in their own testimony. When experts commit the error that Ms. Romero committed, they elevate the risk that jurors will believe that the evidence is stronger than it really is.[44]

---

[40] *Id.* at 461.

[41] *Id.*

[42] *Id.* at 462.

[43] Another noteworthy aspect of the exchange above is defense counsel's objection to the prosecutor's attempt to lure Ms. Romero into approving and committing a source probability error. Defense counsel objects on grounds of "relevance," not misstatement of fact. *Id.* at 461.

He protests that the inversion is not relevant because Romero testified that it's not "scientifically valid." *Id.* However, as the judge correctly notes, Romero did not expressly reject the inversion as scientifically invalid (as she should have). Instead, Romero only expressed an unjustified preference for expressing the RMP in a particular way. *Id.*

[44] Having identified this risk, I should also note that empirical studies with mock jurors frequently show that jurors undervalue DNA evidence relative to Bayesian norms. *See, e.g.*, David H. Kaye et al., *Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Match Probabilities*, 4 J. EMPIRICAL LEGAL STUD. 797, 802 tbl.1 (2007). However, the Bayesian norms generally ignore the role of close relatives and, more importantly, laboratory error. But if one assumes that jurors tend to undervalue DNA evidence, it is possible that source probability errors such as those made by Romero in *McDaniel* may actually *increase* the chance that jurors will give

B. Prosecutor's Fallacy

As egregious as Romero's statistical testimony was, the prosecutor committed an even more serious error in his closing argument when he converted the RMP into a probability that the defendant is guilty. This error, which has been referred to both as the Ultimate Issue Error[45] and, more famously, as the Prosecutor's Fallacy,[46] is committed when the RMP is subtracted from 1 and that value is offered to the jury as the probability that the matchee is guilty as charged. Here is what the prosecutor said:

> Mr. Smith [Prosecutor]: Consider the fact that, what is the percentage that Troy Brown didn't commit this crime? Was it 75 percent? Are you 75 percent sure? Based on the DNA? 90 percent, 99, sometimes people use the phrase, I'm 99 percent sure about that. Well, in this case the evidence shows—how sure can you be? 99.999967 percent sure.[47]

As noted above, the probability that a defendant is innocent or guilty cannot be determined from the RMP alone. If this were true, then no other evidence in the case would be relevant, including evidence pertaining to the defendant's opportunity and ability to commit the crime in question. At best, an extreme DNA RMP can provide strong proof that a particular person is among the small group of people who might be the source of the DNA evidence. But, it does not address the possibility that a person may be the source of the recovered DNA evidence yet not be responsible for the crime charged. However, when jurors are expressly told that the scientific evidence enables jurors to be "99.999967% sure" that the defendant committed a crime, jurors need only determine whether this percentage is sufficiently high

---

DNA evidence the weight that it is deserves.

[45] AITKEN & TARONI, *supra* note 11, at 82; Koehler, *supra* note 23, at 31–32.

[46] *See generally* Thompson & Schumann, *supra* note 21.

[47] Jury Trial Transcript Day 4, September 30, 1994, State v. Brown, No. 5833 (Nev. Dist. Ct. 1994), *reprinted in* 2 Joint Appendix, *supra* note 28, at 588, 730.

to overcome any reasonable doubt they might have about the defendant's guilt. Because few things in life are more than 99.99% certain, some jurors may believe that the statistical evidence in itself provides near certain (and hence sufficient) proof of guilt.

III. THE "NEUTRAL" PRIOR PROBABILITY ASSUMPTION: *GRIFFITH V. STATE*

The inverse errors that arose in *McDaniel* are not unusual in cases involving DNA evidence. A similar set of statistical errors were identified and documented in DNA cases in the early 1990s.[48] Perhaps the most worrisome part about these errors is that they are often expressly defended by experts and courts as mathematically proper. Consider *Griffith v. State*.[49] In *Griffith*, the defendant was charged with raping a profoundly retarded patient at a state psychiatric hospital after the sexually inactive patient became pregnant and had a baby.[50] In support of its charge, the prosecution called the director of a Fort Worth–based DNA laboratory to testify about the statistical significance of a paternity DNA match.[51] The DNA expert planned to present a LR of 14,961 (described as a "paternity index") to describe the significance of the DNA match.[52] He also planned to testify that, by using what he referred to as a "neutral" 0.5 prior probability of paternity, the probability that the putative father was the father of the baby in question was greater than 99.99%.[53]

The defense objected to the introduction of the 0.5 prior probability as well as the computation of a probability of paternity that relied on this prior. I was the defense expert in this case. In a preliminary hearing, I testified that the use of the 0.5 prior probability was neither neutral nor an appropriate matter of speculation for the forensic scientist. Instead, I argued that the

---

[48] Koehler, *supra* note 23, at 28–31.

[49] Griffith v. State, 976 S.W.2d 241 (Tex. Ct. App. 1998).

[50] *Id.* at 242.

[51] *Id.*

[52] *Id.* at 243–44.

[53] *Id.* at 245.

prior probability should reflect the strength of the nongenetic evidence in the case as determined by the finders of fact. I suggested that the academic literature strongly supported my position and that the use of 0.5 as a "neutral assumption" was not generally accepted in the knowledgeable scientific community. Relatedly, I argued that it was inappropriate for the forensic scientist to offer a "probability of paternity" by using Bayes' theorem to combine a 0.5 prior probability with a LR of 14,961. I suggested that the method of using a 0.5 prior probability amounts to an attempt to legitimate an inverse fallacy by turning the LR into a posterior odds ratio. Finally, I suggested that a posterior probability of paternity that is computed in this manner could mislead the jury about the strength and meaning of the genetic evidence.

The trial judge rejected my arguments, admitted the DNA expert's testimony in full, and the defendant was convicted of sexual assault. The verdict was appealed to the Texas Seventh Court of Appeals on the grounds that the 0.5 prior probability violated the defendant's right to be presumed innocent until proven otherwise.[54]

The defense called the court's attention to a 1994 Connecticut Supreme Court opinion, *State v. Skipper*, in which the court rejected Bayesian computations in paternity cases that relied on a 0.5 prior probability.[55] In rejecting *Skipper*, the Texas appellate court mischaracterized *Skipper* as having argued that that the probability of paternity statistic assumes that the putative father did, in fact, have sex with the mother rather than *may have had* sex with the mother. *Skipper* did not rely on this argument.[56] Instead, *Skipper* argued that the introduction of an

---

[54] *Id.* at 242.

[55] State v. Skipper, 637 A.2d 1101, 1107–08 (Conn. 1994).

[56] In fact, the court in *Skipper* noted that:

> [The probability of paternity computation was] predicated on an assumption that there was a fifty–fifty chance that sexual intercourse had occurred in order to prove that sexual intercourse had in fact occurred. The fifty–fifty assumption that sexual intercourse had occurred was not predicated on the evidence in the case but was simply an assumption made by the expert.

Id. at 1106 (citations omitted).

arbitrary 50% prior probability of paternity violated the presumption of innocence.[57] The Texas appellate court ultimately defended the 0.5 prior probability assumption because it is frequently used[58] and "neutral."[59]

The views of the Texas appellate court on the legitimacy of using Bayes' theorem to convert a LR into a posterior odds ratio by assuming a prior of 0.5 are not unique. Earlier this year, another appellate court cited the *Griffith* court's arguments favorably.[60] However, it is far from clear that either of these courts understood the underlying math. Both courts claim that Bayes' theorem is "required" to convert probabilities into percentages.[61] This is not true. As noted earlier, one in 3,000,000 may be described as a probability (.00000033) or as a percentage (.000033%). The conversion of a probability into a percentage is accomplished simply by multiplying the probability by 100 and then placing a "%" at the end of the result. Bayes' theorem has nothing to do with it. Bayes' theorem is a formula that tells decision makers how their prior beliefs about, say, a putative father's paternity, should change in response to new evidence (such as a particular DNA result). It tells decision makers how to move from the probability that a hypothesis is true, to the probability that a hypothesis is true *given new information*.

---

[57] "[W]hen the probability of paternity statistic is introduced, an assumption is required to be made by the jury before it has heard all of the evidence—that there is a quantifiable probability that the defendant committed the crime." *Id.* at 1107–08.

[58] "[M]illions of HLA and DNA tests around the nation reported paternity results using Bayes' Theorem and the probability of paternity invoking a .5 prior probability." *Griffith*, 976 S.W.2d at 246.

[59] "The use of a prior probability of .5 is a neutral assumption. The statistic merely reflects the application of a scientifically accepted mathematical theorem which in turn is an expression of the expert's opinion testimony." *Id.* at 247.

[60] Jessop v. State, 368 S.W.3d 653, 669 n.19, 674 (Tex. Ct. App. 2012).

[61] *Id.* at 669 n.19 ("Bayes' Theorem uses a mathematical formula to determine conditional probabilities and is necessary to convert probabilities into percentages."); *Griffith*, 976 S.W.2d at 243 ("Bayes' Theorem is necessary to convert probabilities into percentages.").

IV. ERROR RATES

The previous two sections documented statistical errors associated with DNA evidence (inverse errors and flawed prior probability assumptions, respectively). One possible response to these errors is to claim that they are inconsequential. According to this argument, even if it is improper to translate a DNA RMP of one in 3,000,000 in *McDaniel* to a source probability of 99.999967%, the extremely small RMP still justifies a strong belief that the matching defendant is, in fact, the source of the recovered DNA evidence. Similarly, even if paternity experts are not justified in assuming a 0.5 prior probability of paternity for all putative fathers, the extreme LRs that are commonly observed in paternity cases should give the fact finder confidence that the putative father is indeed the father of the child in question.

It is true that source probability errors and unjustifiable assumptions about prior probabilities are less significant when RMPs and corresponding LRs are extreme. However, the Prosecutor's Fallacy, wherein the RMP is equated with P(Not Guilty | Match), remains a significant concern when the RMP is extremely small. Even if one infers, from an extremely small RMP, that the matchee is the source of the evidence, this inference should not prompt the additional inference that the matchee must have committed the crime in question. The matchee may be the source of the trace evidence in question, but he or she may not have committed the crime. The trace evidence may have been deposited by the matchee either before or after the crime was committed. Alternatively, the matchee's DNA may have been deposited by the perpetrator himself, either intentionally (as part of a frame up effort) or unintentionally (through inadvertent transfer). In short, those who commit the Prosecutor's Fallacy in cases that include very small RMPs may be relying on weak or irrelevant evidence to justify belief in a defendant's guilt.

In DNA match cases that include very small RMPs, a different consideration should take center stage when gauging the probative value of the evidence: the risk of false positive error.[62]

---

[62] Depending on the facts of the case, the risk that the true source is a

Simply put, the probative value of a DNA match is capped by the frequency with which false positive errors occur.[63] It makes no difference if the RMP is one in millions, billions, or even septillions[64]: if the probability that an analyst will erroneously report a match on two nonmatching DNA samples is 1 in 500, then the corresponding LR is, at best, 500:1.[65] In other words, the false positive error rate—rather than the RMP—tells us most of what we need to know about the probative value of a DNA match. With this in mind, we must ask ourselves whether it pays to risk confusion and various inverse errors by providing fact finders with the RMP at all. Elsewhere I have suggested that in cases where the RMP is several orders of magnitude smaller than the false positive error rate (e.g., RMP = 1 in 1,000,000; false positive error rate = 1 in 500), that the answer is no.[66] There is no need to provide the RMP in such cases because it does not contribute anything beyond the false positive error rate in terms of helping jurors understand a fact in evidence.[67]

What should jurors be told in cases like the one described above? They should be told something like this:

> *The suspect reportedly matches the DNA evidence found at the crime scene. The chance that we would report such a match on nonmatching samples, either because of a coincidence or because of an error, is approximately one in 500.*

---

close relative of the matchee may also be an important consideration. However, as DNA matches are based on more and more loci (currently, about thirteen loci), this risk fades considerably. *See generally* NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *supra* note 5, at 3-12.

[63] *See, e.g.*, Jonathan J. Koehler, *Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter*, 59 HASTINGS L.J. 1077, 1079 (2008) ("[T]he false positive error rate limits and controls the probative value of the match report.").

[64] People v. Odom, No. B225910, 2011 WL 5121175, at *5 (Cal. Ct. App. Oct. 31, 2011) ("[The state's DNA expert] testified that two in 24 septillion people . . . would be expected to match that profile.").

[65] Recall that the LR is approximately the inverse of the RMP (i.e., 1/RMP). *See supra* text accompanying note 14.

[66] Koehler et al., *supra* note 11, at 210.

[67] The Federal Rules of Evidence require that expert testimony be helpful to the trier of fact. FED. R. EVID. 702(a).

Again, the RMP of 1 in 1,000,000 contributes nothing of value beyond this. Indeed, this RMP might actually promote confusion by inviting jurors to commit one of the fallacies described previously. Or it might tempt them to commit other errors such as averaging the RMP with the error rate, or ignoring the error rate altogether based on a mistaken belief that the RMP is the more relevant statistic. Empirical data showing that fact finders are improperly influenced by RMPs in these situations[68] support the argument that introduction of RMPs can be more harmful than beneficial.

At this point, one might wonder whether forensic science statistics of all sorts should simply be hidden from fact finders altogether. Perhaps we should let the forensic scientists handle the numbers in their laboratories but then have those same experts offer more qualitative opinions sans numerical data at trial. The truth is that forensic science testimony rarely includes a quantitative component outside of the DNA context. Non-DNA forensic scientists commonly offer their opinions about who or what is the source of the forensic science evidence (e.g., a hair, a shoeprint, a tire track, a bite mark, a fingerprint, a fiber, etc.). In some domains, forensic scientists use vague terms such as "consistent with," "match," and "could have come from" to explain their failure to find critical differences between two hairs, two fingerprints, etc. The central problem with such terms is that they lack consensus meaning. Two hairs may be "consistent with" one another because they are both brown and thick. Or they may be consistent with one another because they share a large collection of rare features. Without more information about the size of the set of included and excluded features, fact finders may find it hard to assign weight to qualitative terms.

V. LINGUISTIC MESS: PRELIMINARY HEARING

In some forensic areas (e.g., fingerprints and shoeprints), forensic scientists resort to strong language to report their opinions, referring to matches as "identifications" and

---

[68] Koehler et al., *supra* note 11, at 210–11.

"individualizations."[69] Some experts use those two words interchangeably to indicate that the matching person or object is the one and only possible source of the marking to the exclusion of all others in the world. Indeed, phrases such as "to the exclusion of all others in the world" are commonly used by forensic scientists in many non-DNA disciplines to declare their opinions about who or what is the source of an evidentiary item.[70]

Recently, some examiners have tried to distinguish between the words "individualization" and "identification" by suggesting that individualization is a factual state of the world whereas an identification is merely the opinion of the examiner. Consider the following cross-examination of a respected fingerprint examiner in a 2008 preliminary hearing on the admissibility of fingerprint evidence:

> Q: Okay. And by comparing the unknown prints to the known prints, you hope to either declare an individualization or an exclusion between the unknown and the known, correct?
>
> A: Well, when you say individualization and it's kind of a—when I come to my result, I'm actually referring to that as an identification. Individualization, the scientific community, kind of the international, it's ah, more along the lines of excluding it to the possibility of all others on the face of the earth. But when we say an identification . . . . I am telling you that I am confident that that latent print was made by this particular person.
>
> Q: And that is, meaning that particular individual?
>
> A: Yes.
>
> Q: So that would be an individualization; you'd be saying that this individual left that print?
>
> A: Ah, no. . . . [W]hen I say identification, it is my opinion and that I am confident in my result that this

---

[69] MOENSSENS ET AL., *supra* note 4, at 454.

[70] *See* Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199, 206 (2008) (quoting United States v. Green, 405 F. Supp. 2d 104, 107 (D. Mass. 2005)).

latent print and the known prints that I am comparing to were made by the same source.

Q: Meaning that individual?

A: No.[71]

At the beginning of this exchange, the expert distinguishes an individualization from an identification, suggesting that his own identification conclusion is a mere statement about who he believes is the source of the prints, rather than a statement that excludes the possibility that anyone else on earth could be the source. But, in drawing this distinction, the expert appears to directly contradict himself. At first, he says "yes" in response to the attorney's question about whether an identification means that a print was made by "that particular individual?" But seconds later, when the attorney repeats his question ("meaning that individual?"), the expert says "no."[72]

Trial transcripts are littered with confusing exchanges between attorneys and witnesses. Despite this, the exchange above is noteworthy both because the content is important and difficult and because this expert is so highly regarded. Although cross-examination has been referred to as "the greatest legal engine ever invented for the discovery of truth,"[73] cross-examination often does not afford experts the opportunity to expand and clarify answers to complex issues. One can only imagine what the judge (or jury) would take away from the exchange above.

CONCLUSION

What does all of this mean for the burgeoning fields of forensic linguistics and authorship attribution? First, it means that these communities would be wise to set up clear and unambiguous standards for examining materials, documenting their findings, and reporting those findings in court. Doing so

---

[71] Transcript of Proceedings at 48–49, State v. Hull, 788 N.W. 2d 91 (Minn. 2010) (No. 48-CR-07-2336).

[72] *Id.* at 49.

[73] 5 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 1367, at 32 (1974).

will help forensic linguists persuade courts that their evidence is based on reliable methods and will be helpful to jurors.[74] At the very least, expert witnesses should examine materials in a common way, use agreed-upon standards for identifying and recording consistencies and inconsistencies in evidentiary materials, and use a common language to describe findings and conclusions to triers of fact. To facilitate these goals, the forensic linguistics community should establish a professional body that not only promotes these goals but also certifies experts and, where applicable, accredits training programs and laboratories.

As indicated earlier, the forensics linguistics community appears to be divided on the question of whether it favors qualitative versus quantitative methods. Whereas forensic stylists favor the qualitative approach, computational linguists and computer scientists in the field favor a quantitative approach. Regardless of which approach prevails, the field will likely succeed or fail as a function of the scientific quality of its methods. This metric favors the quantitative approach, though the field will need to do a better job developing the requisite databases and transparent methodologies. In an analogous manner, some of the more traditional forensic sciences, such as fingerprinting and voiceprint analysis, are beginning to explore quantitative approaches.[75]

Of course, dangers await. As the field moves toward more probabilistic analyses and outputs, inverse errors may be committed both in and out of the courtroom. It is therefore imperative that the forensic linguistics community identify clear and consistent standards for reporting and testifying about results

---

[74] Scientific evidence must be reliable according to the U.S. Supreme Court. Daubert v. Merrell Dow Pharm., Inc. 509 U.S. 579, 589 (1993). As noted previously, the Federal Rules of Evidence further require that expert testimony be helpful to the trier of fact. FED. R. EVID. 702(a).

[75] Christophe Champod & Ian W. Evett, *A Probabilistic Approach to Fingerprint Evidence*, 51 J. FORENSIC IDENTIFICATION 101, 117–18 (2001); Geoffrey S. Morrison, *Measuring the Validity and Reliability of Forensic Likelihood-Ratio Systems*, 51 SCI. & JUST. 91 (2011) (quantifying the accuracy of forensic voice prints); Cedric Neumann et al., *Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiae*, 52 J. FORENSIC SCI. 54, 54–64 (2007).

and include training in elementary statistics and probability for its members. Regarding the latter recommendation, it is not enough that a forensic field has good scientific intentions and embraces rigorous scientific principles: expert witnesses who provide quantitative testimony must understand enough about statistics and probability to avoid, explain, and correct statistical misstatements when they arise.

The forensic linguistics community should also support a rigorous proficiency-testing program, using realistic evidentiary items, for all techniques and experts. Participation in the program, which should be conducted by an external agency that does not have an interest in demonstrating positive outcomes, should be mandatory for courtroom testimony.[76] Such tests can alert the field and the courts to strengths and weaknesses associated with various techniques and can provide reasonable first-pass estimates for relevant error rates.

Finally, forensic linguistics can learn from the recent battles waged over the individualization claims made by fingerprint examiners.[77] As the exchange in *State v. Hull* documented in Section V indicates, some examiners recognize that individualization claims reach beyond the available data in most

---

[76] Jonathan J. Koehler, *Proficiency Tests to Estimate Error Rates in the Forensic Sciences*, 12 LAW PROBABILITY & RISK 89 (2013); Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 893–94 (2005). Some of the traditional non-DNA forensic sciences appear to be moving in this direction. A recent Expert Working Group report on latent print examination recommended a similar testing program for fingerprint examiners. EXPERT WORKING GROUP ON HUMAN FACTORS IN LATENT PRINT ANALYSIS, LATENT PRINT EXAMINATION AND HUMAN FACTORS: IMPROVING THE PRACTICE THROUGH A SYSTEMS APPROACH 187–88 (David H. Kaye ed., 2012).

[77] *See generally* Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 LAW PROBABILITY & RISK 233 (2009); Simon A. Cole, *Who Speaks for Science? A Response to the National Academy of Sciences Report on Forensic Science*, 9 LAW PROBABILITY & RISK 25 (2010); Jonathan J. Koehler & Michael J. Saks, *Individualization Claims in Forensic Science: Still Unwarranted*, 75 BROOK. L. REV. 1187 (2010); Saks & Koehler, *supra* note 70. *But see* David H. Kaye, *Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies*, 75 BROOK. L. REV. 1163 (2010).

(if not all) forensic sciences. Forensic linguistics would do well to *offer conservative, descriptive claims and to support those claims with empirical data*. Source claims (e.g., "In my opinion, this text was written by the defendant") should be avoided. Such a modest approach will not only help forensic linguistics gain a place in the courtroom, but it will also reduce the risk that jurors will overweigh this potentially important, but as yet untested, evidence.