ERROR AND EXAGGERATION IN THE PRESENTATION OF DNA EVIDENCE AT TRIAL

Jonathan J. Koehler*

DNA identification evidence has been and will continue to be powerful evidence against criminal defendants. This is as it should be. In general, when blood, semen or hair that reportedly matches[1] that of a defendant is found on or about a victim of violent crime, one's belief that the defendant committed the crime should increase, based on the following chain of reasoning:

Match Report → True Match → Source → Perpetrator

First, a reported match is highly suggestive of a true match, although the two are not the same. Errors in the DNA typing process may occur, leading to a false match report. Second, a true DNA match usually provides strong evidence that the suspect who matches is indeed the source of the trace, although the match may be coincidental. Finally, a suspect who actually is the source of the

*Jonathan J. Koehler is Assistant Professor, Management Science and Information Systems Department, College of Business Administration, University of Texas at Austin. Research on this Article was supported by National Science Foundation grant SES-9209544. The research assistance of Sam Lindsey and Audrey Chia and the constructive criticisms made by Richard Lempert and an anonymous reviewer are gratefully acknowledged. A version of this paper was presented at the Second International Conference on Forensic Statistics, Tempe, Arizona, March 1993.

1. Some have argued that "match" versus "no match" terminology imposes an arbitrary dichotomy and obscures the diagnostic value of the test results. BERNARD ROBERTSON & TONY VIGNAUX, UNDERSTANDING EXPERT EVIDENCE (forthcoming). However, I believe that the match classification is useful provided that pains are taken to explain that the diagnostic value of matches varies greatly. The alternatives that have been proposed (*e.g.*, provide jurors with Bayesian likelihood ratios) have technical merit, but they may be difficult to understand. Such questions about the efficacy of different modes of presentation can and should be addressed empirically.

trace may not be the perpetrator of the crime. The suspect may have left the trace innocently either before or after the crime was committed.

In general, the concerns that arise at each phase of the chain of inferences are cumulative. Thus, the degree of confidence one has that a suspect is the source of a recovered trace following a match report should be somewhat less than one's confidence that the reported match is a true match. Likewise, one's confidence that a suspect is the perpetrator of a crime should be less than one's confidence that the suspect is the source of the trace.

Unfortunately, many experts and attorneys not only fail to see the cumulative nature of the problems that can occur when moving along the inferential chain, but they frequently confuse the probabilistic estimates that are reached at one stage with estimates of the others. In many cases, the resulting misrepresentation and misinterpretation of these estimates lead to exaggerated expressions about the strength and implications of the DNA evidence. These exaggerations may have a significant impact on verdicts, possibly leading to convictions where acquittals might have been obtained.

This Article identifies some of the subtle, but common, exaggerations that have occurred at trial, and classifies each in relation to the three questions that are suggested by the chain of reasoning sketched above: (1) Is a reported match a true match? (2) Is the suspect the source of the trace? (3) Is the suspect the perpetrator of the crime? Part I addresses the first question and discusses ways of defining and estimating the false positive error rates at DNA laboratories. Parts II and III address the second and third questions, respectively. These sections introduce the "source probability error" and "ultimate issue error" and show how experts often commit these errors at trial with assistance from attorneys on *both* sides. Part IV introduces two related exaggerations, the "P(Another Match) error" and the "numerical conversion error." Part V provides a simple and general explanation for the persistence the errors identified. Part VI concludes with a discussion of the ways in which scientists can take advantage of their roles as teachers, expert witnesses, and researchers to educate the courts about the meaning and limits of probabilistic DNA evidence.

I. IS THE REPORTED MATCH A TRUE MATCH?

A reported match is a true match if the characteristics that are identified by the analysis as belonging both to the trace and to the suspect's sample are, in fact, characteristics of the trace and the suspect. Forensic scientists are often reluctant to acknowledge that a reported match could be something other than a true match. When asked about the possibility, many respond by discussing their skill, care and experience in typing samples, or by discussing the validity of their protocol. Such discussion provides only a very limited basis for assessing the probability that a reported match is a true match. While it may be true that laboratories that have superior procedures are less likely to commit

errors, the primary concern is not with the process that yields conclusions but with the accuracy of the conclusions themselves.

What can go wrong? First, technical errors are possible. According to testimony provided by Dr. Robert Kidd in *People v. Axell*,[2] enzyme failures, abnormal salt concentrations, and mischievous dirt spots can produce misleading DNA banding patterns. Human errors are also possible. Contaminations, mislabelings, misrecordings, misrepresentations, case mix-ups and interpretive errors may lead to false positive errors.

Some of these errors have been documented in proficiency tests as well as in actual casework.[3] However, many forensic scientists who testify in court are reluctant to acknowledge even the *possibility* of false positive error: "[I]t is technically impossible to make a false/positive identification."[4] "There is no way to get a false positive with this technology."[5] An incorrect match is an "impossible" result.[6] DNA analysis is "failsafe."[7] The accuracy rate is 100%.[8] And so on.[9]

---

2. People v. Axell, 235 Cal. App. 3d 836, 1 Cal. Rptr. 2d 411 (1991).

3. CALIFORNIA ASS'N OF CRIME LABORATORY DIRECTORS, DNA COMMITTEE REPORT NO. 6 (Oct. 1, 1988) (hereinafter CACLD I); Simon Ford & William C. Thompson, *A Question of Identity: Some Reasonable Doubts About DNA "Fingerprints,"* 30 SCIENCES 37 (Jan.–Feb. 1990); Eric Lander, *DNA Fingerprinting on Trial*, 339 NATURE 501 (1989); Thompson & Ford, *The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints*, *in* FORENSIC DNA TECHNOLOGY (Farley & Harrington eds., 1991).

4. Jones v. State, 569 So. 2d 1234 (Fla. 1990) (transcript at 677).

5. Kelly v. State, 792 S.W.2d 579 (Tex. App. 1990), (transcript at 919); *see also* Yelder v. State (Ala. Crim. App. 1991) ("there's no way to make or create a false positive with this test," transcript at 84); State v. Pierce, No. 89-CA-30 (Ohio App. 1990) ("You can't get a false positive," transcript at 431).

6. Cobey v. State, 80 Md. App. 31, 559 A.2d 391 (1989).

7. People v. Fishback, 829 P.2d 489 (Colo. App. 1991), *aff'd*, 851 P.2d 884 (Colo. 1993).

8. State v. Davis, 814 S.W.2d 593 (Mo. 1991) (transcript at 82); *see also* Hicks v. State, No. 70,803 (Tex. 1993) ("According to Caskey, a false positive was impossible because if the procedures were not correctly followed, no match could be obtained."); Kelly v. State, 792 S.W.2d 579 (Tex. App. 1990) ("a false 'match' of a known DNA sample with an unknown DNA sample is impossible with RFLP technique," transcript at 570); People v. Wesley, 183 App. Div. 2d 75, 589 N.Y.S.2d 197 (1992) ("it was unrefuted that it is impossible under the RFLP procedure to obtain a false positive result"); State v. Harris, No. 01-C-01-9010-CR-00258 (Tenn. Crim. App. 1992) ("Agent Adams testified that . . . it would be impossible to get a false DNA match," transcript at 11). *But see* State v. Alt, K4-90-1437 (Minn. App. 1993) ("Certainly the RFLP process lends itself to error," Order Limiting the Use of DNA Test Results, May 29, 1992, at 3).

9. Consider the following exchange in *State v. Bethune*, 821 S.W. 2d 222 (Tex. App. 1991) (transcript at 2228):

> Q: Now, you're telling us that you can only get a result or no result; is that correct?
> A: That's correct.
> Q: And you couldn't get a false positive?
> A: There's nothing like a false positive in this, no.
> Q: How about if you use the wrong sample?
> A: If you use the wrong sample?
> Q: (Nods head)
> A: You either get a result, or you don't get a result. There's no false positives.

These statements are extremely misleading and may be reversible error. A factfinder needs to know how likely it is that a reported match is not a true match. This probability, in combination with the probability that a suspect who truly matches a trace is not its source, comprises the false positive error rate. Notice that technical errors and human performance errors contribute to the false positive error rate.

Some forensic scientists will object to this definition of false positive error. They prefer a narrower definition that includes only those errors that result from *technical* failures. With this definition in mind, some say, the impossibility statements above are justifiable.

Two points should be made in response. First, technical errors are not only possible, they have occurred in some instances. For this reason alone, the impossibility claims should be forbidden. But even assuming that the probability of technical error is negligible, experts should not be permitted to equate the technical error rate with the false positive error rate. Judges and jurors are (or should be) concerned with identifying the rate at which false positive errors occur for whatever reason, rather than the rate at which false positive errors arise for a particular reason. Those who insist on defining false positive error as error that arises in a particular way are engaged in a sinister semantic game. There is a danger that jurors will understand their testimony to refer to the likelihood of false identification rather than the likelihood of a certain type of false positive error. This is particularly likely in cases such as *Bethune* where the expert repeated his impossibility claim even after being asked about the possibility of a human error.[10]

The best way to measure the rate of false positive error associated with a laboratory or an individual technician is through an ongoing series of blind, external proficiency tests conducted under realistic conditions. In these tests, samples of genetic materials such as blood, semen and hair can be provided to laboratory technicians who are then asked to determine which, if any, match samples taken from possible sources. Erroneous match reports between "recovered" samples and suspected sources constitute false positive error. Failures to detect matches between recovered samples and true sources constitute false negative error. Further investigation of these errors may reveal their causes and lead to procedural modifications and improved performance.

Surprisingly, there have been no blind external proficiency tests conducted to date. In the few tests conducted by outside agencies, the tested laboratories and technicians were aware that they were being tested. This makes inferences from test performance to case work performance difficult. It may be, for example, that the technicians who conduct DNA analyses are more diligent and cautious when they know that they are being observed and tested.

10. *See supra* note 9.

An equally serious problem is that most proficiency tests have not used samples that are representative of casework. Sample stains in tests are usually large and carefully preserved on a clean cotton cloth. Moreover, the laboratories are often told what the samples are composed of and how and when they were prepared.

Although many DNA laboratories prefer to conduct in-house proficiency tests rather than submit to external testing, some outside tests have been conducted.[11] Tests conducted by the California Association of Crime Laboratory Directors (CACLD) on three DNA laboratories in 1987 and 1988 revealed several false positives. In an initial study, 50 samples were sent to each of the three laboratories. Two of the laboratories reported their results in match-no match terms, while the third laboratory reported its results in terms of the DQ-alpha genotype. Sixty-six matches were reported by the first two laboratories; one match was a false positive. The third laboratory typed 47 samples and misclassified one.[12] In a follow-up study based on 50 samples, the first two laboratories reported 91 matches, of which one was a false positive.[13]

More recently, Collaborative Testing Services (CTS) conducted proficiency tests on 38 DNA laboratories.[14] These tests offer insight into the frequency and types of errors that occur in DNA analyses under extremely favorable test conditions.[15] Although CTS concluded that "there were no false matches," an analysis of the report suggests otherwise. I found at least three false positive errors out of an estimated 75 match reports.[16]

---

11. One recent external proficiency test was sent to 94 laboratories, 60 of which did not return data. COLLABORATIVE TESTING SERVICES, DNA PROFILING, REPORT NO. 91-15 (1992) [hereinafter CTS].

12. CACLD I, *supra* note 3. Although not a false positive in the strict sense of the term, a misclassification can easily lead to a false positive error in the DQ_α system, which has a limited number of possible phenotypes.

13. CALIFORNIA ASSOCIATION OF CRIME LABORATORY DIRECTORS, DNA COMMITTEE, RESULTS OF BLIND TRIAL NO. 2 (Mar. 29, 1990) (hereinafter CACLD II).

14. CTS, *supra* note 11. The American Association of Blood Banks also has conducted DNA proficiency tests, but these did not score the performance of participating laboratories.

15. By favorable conditions, I have many aspects of the tests in mind. First, participation in the tests was voluntary. Second, the tests were nonblind. Third, the labs were told how many stains there were and whether or not the stains were mixed. The detection of mixed stains in a sample is itself an important indicator of laboratory quality, and recently played a key role in a widely publicized case. *E.g.*, John F. Harris & Robert F. Howe, *Wilder to Allow Execution: Barring Court Reprieve Coleman Dies Tomorrow in VA's Electric Chair*, WASHINGTON POST, May 19, 1992, at 1. Fourth, the samples were larger than normal "in order that the 'new' mixture was not clouded by a quantity element." CTS, *supra* note 11, at 1. Obviously, stain quantity plays an important role in actual casework in that small quantities may not lend themselves to re-analysis. Indeed, in response to criticism that the samples were unrealistic, CTS admitted that the mixed sample "was not truly representative of case samples. The sample was . . . prepared by 'dunking' and not by 'swabbing.' " *Id.* Finally, one of the samples used in the test was identical to a sample used in an earlier CTS test. This raises the possibility of sample recognition.

16. In the CTS study, laboratories were presented with five blood samples (A-E) and one blood/semen mixture sample (F). Samples A-E were taken from five different people; sample F consisted of blood from the female source of sample A and semen from the male source of sample D. After correctly concluding that the five bloodstains A-E came from different individuals,

Taken together, all these results suggest that false positive errors occur in one to four percent of match reports provided in open proficiency tests. Although it is hard to say whether the false positive error rate in actual casework is much different than this, it is clear that reported matches are not always true matches.[17]

II. IS THE SUSPECT THE SOURCE OF THE TRACE?

A. The Reference Population

Even if a reported match is a true match, the suspect will not be the source of the trace if the match is purely coincidental. To the extent that the frequency of the matching traits, F(Traits), is rare, the probability that the suspect actually is the source increases.[18] But before F(Traits) may be estimated, some attention must be given to identifying an acceptable reference population.

The reference population used by forensic science laboratories to derive F(Traits) is usually based on the ethnic group of the suspect (e.g., black, Hispanic, Caucasian).[19] Though convenient, this practice is misguided. It is only appropriate when it is known that the source of the recovered trace is a member of the suspect's ethnic group.[20] When there is *no* information about the ethnic group of the trace source, the general population is a more appropriate reference class. When there is *some* information about the trace source, it would be best to compute F(Traits) based on a case-specific "potential source population."[21]

---

laboratory 1504 falsely reported that "the profiles from sample D matched those obtained from the female fraction/blood on the swab [sample F]. The profiles from sample A matched those obtained from the male fraction." Laboratory 1518 concluded that "sample D matched the DNA profiles from sample F, male fraction and sample F, female fraction 2 with all four probes used." Although their band size data do not support this mistaken conclusion, this sentence appears to suggest that both the male and female portions of F matched sample D. Laboratory 1528 falsely concluded that samples B and E had the same origin. This was a coincidental match based on a single probe. Laboratory 1532 erroneously matched the female blood portion of F with sample D on each of three probes.

17. The false positive error rate may actually be lower in casework if (a) the laboratories that do the bulk of the analyses are less likely to commit false positive errors than the average laboratory in the tests, (b) the base rate frequency for true matches is higher in casework than in these proficiency tests.

18. F(Traits)=The random match probability—the probability that a person selected at random from the reference population will match the trace evidence.

19. In many cases, frequency estimates based on more than one of these broad classes are provided.

20. People v. Pizarro, 10 Cal. App. 4th 57; 12 Cal. Rptr. 2d 436 (1992) ("the relevancy of the statistical probability depends on the perpetrator being the same racial or ethnic background as the suspect. . . . Absent proof to support the preliminary fact as to the racial/ethnic background of the perpetrator, we see no relevancy to a database selected because of the racial/ethnic background of the suspect/defendant").

21. Others have argued that the ethnic group of the suspect should be rejected as a reference population in favor of the "suspect population," of potential perpetrators of the crime. Richard Lempert, *Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to*

Practical problems arise in the construction of potential source populations. These include the need to construct them on a case by case basis, and the lack of clear standards for deciding who is and is not a member of the population. When disagreements arise about the composition of the population, F(Traits) can be computed for several different source populations. In many cases, the resultant F(Traits) values will be sufficiently similar that there is no practical effect for using one population rather than another. But in other cases—particularly those in which there is disagreement about whether particular relatives of the suspect should be included in the source population—the differences may be important.

In general, the inclusion of a suspect's close relatives in the potential source population will lead to F(Traits) values that are larger—hence less diagnostic—than F(Traits) values constructed on the general population. This is because a suspect's relatives are more likely to be genetically similar to him or her than a random member of the general population. F(Traits) based on three probes may be one in millions for the general population. But the probability that the suspect and his biological brother will share a set of alleles on each of the three probe sites is approximately $(1/4)^3 = 1/64$.[22]

B. The Source Probability Error

Even in cases where there is no dispute about F(Traits), there may be confusion about its significance for estimating the probability that the suspect is the source of the matching trace. Specifically, there is a tendency to equate F(Traits), with the probability that the suspect is not the source of the trace, P(Not-Source). Equating F(Traits) with P(Not-Source) tends to exaggerate the strength of the DNA evidence, and may be referred to as the "source probability error." Absent an estimate of the size of the potential source population, a source probability statement cannot be made. A Bayesian analysis illustrates the point. Bayes's Theorem states that the odds on the defendant being the source given the reported match are the prior odds on this hypothesis times the likelihood ratio

---

the Reverend Bayes, 13 CARDOZO L. REV. 303, 310 (1991). *See also* David A. Stoney, *Reporting of Highly Individual Genetic Typing Results: A Practical Approach*, 37 J. FORENSIC SCI. 373, 380 (1992). In many cases, the suspect population approximates the potential source population. But there may be members of the potential source population who are excluded from the suspect population, and vice versa. Imagine, for example, a case in which a woman is murdered in her bed one week after her husband died. Hairs recovered at the scene of the crime may belong to the woman's deceased husband. This places the woman's husband in the potential source population, although he would not be a member of the suspect population. Or, consider the possibility outlined in Scott Turow's best selling novel *Presumed Innocent*: a woman commits a murder and subsequently plants her husband's semen in the victim in an effort to incriminate him. Here, the woman might be a member of the suspect population for reasons that have nothing to do with the trace evidence, but she would not be a member of the potential source population.

22. Jonathan J. Koehler, *DNA Matches and Statistics: Important Questions, Surprising Answers*, 76 JUDICATURE 222 (1993).

L = P(Reported Match | Source)/P(Reported Match | Not-Source). While L is not connected to the size of the source population, the prior odds P (Source)/ P(Not-Source) are. Absent other background information, if the source population consists of only ten people, of which the defendant is one, then the prior odds are one to nine. If the source population consists of one million people equally likely to be the source, the prior odds are one to 999,999. In this way, the size of the population informs one's estimate of the prior odds. Since the prior odds affect the posterior odds, estimates of P(Source | Reported Match) cannot be made on the basis of forensic identification evidence alone.

Source probability errors are frequently committed in the popular press.[23] They are also committed by the courts and by experts who should know better.[24] After testifying that a DNA match was found between blood from a murder

23. Larry Still, *Genetic Fingerprinting Pointed to a Killer*, VANCOUVER SUN, Feb. 1, 1993, at 1A ("Subsequent DNA comparisons not only showed blood on the suspect's jeans came from the victim, but also proved a semen sample taken from the victim's vagina came from the suspect.").

24. United States v. Jakobetz, 955 F. 2d 786, 789 (2d Cir. 1992) ("The FBI . . . calculated that there was one chance in 300 million that the DNA from the semen sample could have come from someone in the caucasian population other than Jakobetz."); United States v. Martinez, No. 91-1996 (8th Cir. 1993) ("The FBI concluded that there was a 1 in 2600 probability that the semen found on the panties came from someone other than Martinez."); People v. Axell, 235 Cal. App. 3d 836, 844, 1 Cal. Rptr. 2d 411 (1991) ("that the frequency of that DNA banding pattern in the Hispanic population is approximately 1 in 6 billion . . . meant that the chance that anyone else but appellant left the unknown hairs at the scene of the crime is 6 billion to 1"); People v. Lindsey, No. 90CA0556 (Colo. App. 1993) ("A genetic epidemiologist . . . testified . . . that the odds of someone besides the defendant having the banding pattern appearing in the known sample and in the forensic sample was one in three hundred forty billion"); People v. Fishback, 829 P.2d 489, 492 (Colo. App. 1991) ("an expert in population genetics and population biology testified that there was a probability of only one in 830,000,000 that someone other than the defendant would match the DNA found in the samples from the victim"); *aff'd*, 851 P.2d 884 (Colo. 1993); People v. Stanley, 246 Ill. App. 3d 393, 615 N.E.2d 1352 (1993) ("the chances of the blood containing all three characteristics belonging to anyone other than the victim are only 1 in 37,500"); People v. Miles, 217 Ill. App. 3d 393, 404; 577 N.E.2d 477, 484 (1991) ("the suspect's DNA is compared to information in Cellmark's African-American data base to determine the probability of an African-American other than the defendant leaving the semen stain on the bed sheet. Foreman testified the probability was 1 in 300,000."); Smith v. Deppish, 248 Kan. 217, 221, 807 P.2d 144, 148 (1991) ("According to the State's three experts, there was more than a 99 percent probability that Smith was a contributor of the semen found on the swab."); Polk v. State, 612 So. 2d 381 (Miss. 1992) ("in this case, the probability that the blood on the waistband of Polk's underwear was from any person other than Georgia Mae Thomas was calculated to be 1 in 530,000,000"); State v. Lee, No. 90CA004741 (Ohio App. 1990) ("The expert concluded that the probability that the DNA found in the semen samples came from anyone other than appellant was one in seven million."); Glover v. State, 825 S.W.2d 127, 128 (Tex. Crim. App. 1992) ("the jury heard expert testimony that the odds were one in eighteen billion that the DNA contained in the vaginal swab specimens taken from the victim belonged to someone other than the defendant"); Kelly v. State, 792 S.W.2d 579, 582 (Tex. App. 1990) ("The statistical probability that the semen came from another white male was 1 in 13.5 million."), *aff'd*, 824 S.W.2d 568 (Tex. Crim. App. 1992); Spencer v. Commonwealth, 238 Va. 275, 289, 384 S.E.2d 775, 782 (1989) ("the chance that anyone other than Spencer produced the semen stains was one in 135 million"). *See also* William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?* 52 LAW & CONTEMP. PROB. 9, 20–21 (1989) (related observation with of other types of forensic evidence).

victim and blood recovered from a blanket, an FBI scientist in *Wike v. State*[25] was questioned by a prosecuting attorney as follows:

> Q: And in your profession and in the scientific field when you say match what do you mean?
> A: They are identical.
> Q: So the blood on the blanket can you say that it came from Sayeh Rivazfar [the victim]?
> A: With great certainty I can say that those two DNA samples match and they are identical. And with population statistics we can derive a probability of it being anyone other than that victim.
> Q: What is that probability in this case?
> A: In this case that probability is that it is one in 7 million chances that it could be anyone other than the victim.

As we have seen, the expert's claim that population statistics alone enable him to determine the probability that the victim is not the source of the recovered blood trace is false.[26] The Florda Supreme Court in *Wike* interpreted the source probability hyperbole to mean that the blood on the blanket was "positively" identified as belonging to the victim.[27]

Such exaggerations of scientific evidence and testimony are common and troubling in their own right. Many judges are quick to assume that reported DNA matches are dispositive of identity even when such conclusions are *not* expressed by the scientific experts.[28] On occasion, the experts *do* testify that they are 100% certain that a particular trace came from a particular person.[29] From a normative standpoint, such testimony is more egregious than the source probability error because it does not even allow for the *possibility* that someone other than the matchee is the source of the trace.[30] Even if source probability

25. 596 So. 2d 1020 (Fla. 1992) (transcript at 417–18).

26. It bears repeating that population statistics are uninformative with respect to the prior odds ratio P(Source)/P(Not-Source); hence, the posterior odds ratio P(Source | Reported Match)/P(Not-Source | Reported Match) cannot be identified. Of course, the expert may be able to identify a range of posterior odds ratios that correspond with various prior odds ratios. But the expert has no special expertise in identifying prior probabilities, hence he or she should not offer an opinion about which ratio is most nearly accurate.

27. *Id.* at 1022.

28. *See* State v. Blair, 592 N.E.2d 854 (Ohio 1990); Hopkins v. State, 579 N.E.2d 1297 (Ind. 1991); State v. Wimberly, 467 N.W.2d 499 (S.D. 1991).

29. *E.g.*, State v. Cauthron, 846 P.2d 502, 516 (Wash. 1993) (experts testified that the defendant "is the source of the semen sample in the five cases that we got the result on," that she had "no doubts about the identification," that the DNA could not have come from anyone else on earth, and that the defendant "was the donor of the semen in those five cases").

30. From an empirical standpoint, this "source certainty error" may not lead to greater verdict bias against defendants than the source probability error. First, people may treat an extremely high source probability statement as if it were a source certainty statement. Second, a source certainty statement may be regarded as less credible than a source probability statement because it fails to admit the possibility of error. Furthermore, if the implicit error bars that people assign to statements of certainty are larger than those assigned to very high and seemingly precise source probability statements, source certainty statements may have *less* impact on a juror's beliefs and verdicts than very high source probability statements.

estimates could be made, it is not clear that a forensic scientist should offer personal interpretations, let alone ones that further exaggerate the strength of the DNA evidence.

C. With "Help" From Attorneys

Trial transcripts reveal that courtroom source probability errors are usually committed with the help of statements in the form of questions from attorneys. Consider the following exchange:

> Q: In layman's terms, just so I get this right, are you saying that the probability that the DNA that was found in the question samples came from anyone else besides Amos Lee is one in 7,000,000, it came from another unrelated person other than Amos Lee?
> A: Yes, approximately.[31]

Even if F(Traits) were indeed one in seven million, the expert only could say that there is one chance in seven million that a single randomly selected person would match the trace evidence. This is not equivalent to a claim that there is one chance in seven million that the DNA came from someone other than Lee. This would only be true in the special case where Lee was one of two equally likely members of the potential source population. If the potential source population contained more than two people (as it usually does), then the probability that the DNA came from someone other than Lee would be greater.

In some cases, the source probability error is committed in the context of a longer attorney/expert exchange that makes it difficult to catch and correct:

> Q: And are you able to compile all four of those probabilities and determine what is the likelihood of the DNA found in Billy Glover just randomly occurring in some other DNA sample?
> A: Yes.
> Q: What is the likelihood of that?
> A: The way that is done is to multiply each one of those four numbers that I mentioned before together, because each one is separate and independent, and the final number comes out as one in about 18 billion.
> Q: So the likelihood that DNA belongs to someone other than Billy Glover is one in 18 billion?
> A: That is correct.[32]

The expert was initially asked about F(Traits). But the attorney rede-

---

31. State v. Lee, No. 90CA004741 (Ohio App. Dec. 5, 1990) (transcript at 464); *see also* State v. Pierce No. 89-CA-30 (Ohio App. 1990) (transcript at 498):

> Q: [S]ome numbers have been given to me which I assume came from Cellmark. They were either done by yourself or somebody there, but in that particular case it is 40 billion to one, that this match had to be Louis Pierce; is that correct?
> A: Yes, that's true.

32. State v. Glover, 825 S.W.2d 127 (Tex. Crim. App. 1992) (transcript at 413).

scribed this value as a source probability, and the expert confirmed this characterization.[33]

The conversational dynamic that exists between attorneys and experts during direct and cross-examinations may be partially responsible for some source probability errors, including those committed overtly by the experts. In a Missouri case, an expert testified that trace evidence had the "same blood types and the same DNA profile as Mr. Davis."[34] But when the prosecuting attorney restated this testimony as "the staining on the lower part of the jacket that you identified as Jack Davis's blood," the expert made no effort to correct this subtle distortion. Likewise, when the expert stated that a particular blood stain was "consistent with Mr. Davis's," the prosecuting attorney interrupted to asked "Which one consists of Mr. Davis's?" Rather than explain that there is an important difference between blood that is consistent with Mr. Davis's and blood that *is* Mr. Davis's, the expert simply answered the misleading question. Eventually, the persistent mischaracterization of the expert's testimony by the prosecuting attorney broke down the expert's scientific venire:

> Q: [W]hose blood was found to be on item 52?
> A: Mr. Davis's blood.[35]

The experts are at least partially to blame for committing and confirming source probability errors. They should know enough about the meaning of a match to resist characterizing F(Traits) evidence in P(Source) or P(Not-Source) terms. They certainly should know enough to avoid absolute identity claims. On the other hand, under the stress of direct and cross-examinations, it may not be reasonable to expect an expert to correct all subtle distortions and misunderstandings expressed by attorneys and judges.

III. IS THE SUSPECT THE PERPETRATOR OF THE CRIME?

An error related to, but more egregious than, the source probability error occurs when F(Traits) is identified as P(Not-Guilty), the probability that the

---

33. This pattern is evident in other cases as well. In a Missouri case, a DNA expert provided F(Traits) testimony by indicating the probability that the various traits would occur in an individual. The following exchange took place on cross examination:

> Q: Well, the probability of it occurring in an individual is to say the probability of this being the person?
> A: If I understand your question correctly, I guess it's a correct conclusion.

State v. Thomas, 830 S.W.2d 546 (Mo. App. 1992) (transcript at 588). Here, too, an attorney committed the source probability error, and the expert failed to correct it. A different expert in the case committed the error more overtly: "Our conclusion can be stated in some other way, which is that although this is not an absolute identification, the likelihood that this sample derived from this particular individual is somewhere in the range of higher than 99.99 percent probability." *Id.* at 600.

34. State v. Davis, 814 S.W.2d 593 (Mo. 1991) (transcript at 2112).

35. *Id.* at 2113, 2123, 2127.

defendant is not guilty. Dubbed the "Prosecutor's Fallacy,"[36] its commission by experts and defense attorneys justifies a broader and more descriptive phrase. Because it mistakes F(Traits) for a probability statement about the ultimate issue, the error of presuming that F(Traits) = P(Not-Guilty) is referred to here as the "ultimate issue error."

People v. Collins[37] is the most famous illustration. In *Collins*, the prosecutor obtained a robbery conviction against a couple by equating the probability that a random couple would possess a series of observed characteristics[38] with the probability that the accused couple did not commit the robbery. The California Supreme Court overturned this conviction, identified errors in the prosecution's assumptions and probabilistic logic, and delivered a stern warning about the dangers of "trial by mathematics." This case has been analyzed extensively and the legal community appeared to understand the difference between probability evidence and the probability of the ultimate issue.

But the ultimate issue error has resurfaced with alarming frequency in cases—particularly rape cases—involving DNA evidence. Some have suggested that the discovery of a DNA match between a defendant and semen recovered from a rape victim justifies an assertion that the probability that someone other than the defendant committed the rape equals F(Traits). In a Texas case, the following exchange took place after the statistics in a DNA report were reviewed:

> Q: Is that correct? So that in the event that the accused sitting in this chair would happen to be White, you're telling the members of this jury that there would [be] a one in 5 billion chance that anybody else could have committed the crime; is that correct?
>
> A: One in 5 billion, correct.[39]

As with source probability errors, judges sometimes commit ultimate issue errors even when the experts do not.[40] Direct evidence that jurors who hear F(Traits) testimony commit ultimate issue errors is harder to come by. But if popular press accounts of DNA testimony indicate or influence how this evidence will be interpreted, ultimate issue errors may be common.[41]

---

36. 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 (1987).

37. 68 Cal. 2d 319, 66 Cal. Rptr. 497, 438 P.2d 33 (1968).

38. The traits were a white woman with blonde hair and a pony tail riding in a partially yellow car with a black man who had a beard and a mustache.

39. State v. Bethune, 821 S.W.2d 222 (Tex. App. 1991) (transcript at 2327).

40. Jones v. State, 569 So. 2d 1234 (Fla. 1990); McElroy v. State, 592 N.E.2d 726 (Ind. App. 1992); Ross v. State, No. B14-90-00659 (Tex. App. 1992).

41. In some press accounts, the error is strongly implied: Susan Warren, *DNA "Fingerprints" May Identify Rapist*, HOUSTON CHRONICLE, Jan. 23, 1988, at 1, 12 ("If the DNA matches, police know they have the rapist."). Other times, it is committed explicitly. Lori Montgomery, *DNA Test Accuracy On Trial: Method is Subject to Error, Critics Say*, DALLAS TIMES, Oct. 14, 1990, at A1, A12 ("The odds that anyone else raped 14-year old Danielle Lemieux and fatally stabbed her, her younger sister and a family friend, the jury was told, were 1 in 54 billion."); B. Merrifield, *Microbi-*

IV. RELATED EXAGGERATIONS

A. The P(Another Match) Error

The mistaken belief that F(Traits) is identical to the probability that there exists another person who matches the defendant's DNA profile, P(Another-Match), is a close cousin of the previous errors. In a Virginia case, an attorney incorrectly restated the expert's F(Traits) testimony:

> Q: I guess I don't understand. You have told the ladies and gentlemen of the jury that the odds are 705 million to 1 against two persons having the pattern that Spencer has; is that correct?
>
> A: That's correct.[42]

The problem here is especially subtle. On the one hand, it is true that for F(Traits) = one in 705 million, there is one chance in 705 million that a single randomly selected member of the reference population will match the observed trait pattern. But the chance that *some* member of the reference population will match the observed pattern may be much greater.

To determine P(Another Match), an estimate of the size of the potential source population must be made. For populations of size N in which F(Traits) = 1/X,

$$P(\text{Another Match}) = 1-(1-1/X)^{N}.^{43}$$

Thus, if F(Traits) = one in 705 million and the potential source population consists of one million unrelated people,

$$P(\text{Another Match}) = 1-(1-1/705{,}000{,}000)^{1{,}000{,}000} = .14\%.^{44}$$

Although P(Another Match) estimates are commonly provided in cases involving DNA evidence, this computation is never made. Moreover, there

---

ologist Challenges Results of FBI DNA Tests in Moore Case, CHI. TRIB. (Zone D), June 9, 1992, at 3 ("The FBI testimony put the odds at 466-1 against someone other than Moore having been responsible for the rape of Zeman."); Mark Platt, *S.D. Officer Given 56 Years in Beach Attacks*, L.A. TIMES, Aug. 11, 1992, at A1 ("Law enforcement officials said the probability that someone other than Hubbard had raped the women ranged from 1 in 340,000 to 1 in 7.7 billion."); *Murder, Rape and DNA*, NOVA, aired Mar. 2, 1993 ("Unlike the PCR method RFLP [a DNA technique] can actually identify the perpetrator of a crime."). Such statements may very well persuade jurors and future jurors to treat F(Traits) evidence as a proxy for P(Not-Guilty).

42. Spencer v. Commonwealth, 238 Va. 295, 384 S.E.2d 785 (1989) (transcript at 557).

43. Purists may wish to be careful about using an N that represents the population of possible sources that excludes the suspect. But in practice, it will make little difference whether the suspect is included or not except when the potential source population is very small and the trait frequency is unusually large.

44. When F(Traits) is larger, P(Another Match) values may be quite large as well. Thus, if F(Traits) were 1/7,000,000, then P(Another Match) would be 13.3%. P(Another Match) values would be even larger if close relatives of the suspect were among those in this potential source population.

appears to be little awareness among experts, attorneys or judges that the size of the potential source population is relevant, let alone necessary, to estimate P(Another Match).

B. The Numerical Conversion Error

Sometimes DNA experts describe the significance of F(Traits) in terms of the number of people who would have to be tested before one should expect another match to occur. This computation is straight-forward, although it is not, as some have said, the denominator of F(Traits). A conclusion that there is one chance in 100 that a randomly selected individual would match as well as the defendant is *not* equivalent to a conclusion that 100 people would need to be tested before another match might be expected. This common mistake may be called the "numerical conversion error."

In a Texas case, the DNA expert was questioned about the F(Traits)=one in 23 million statistic he provided:

> Q: Could you explain briefly to the jury what 1 in 23 million means in reference to this case? What does that mean?
>
> A: It means that we calculated a match for four probes and that the pattern for the suspect in this case occurs in 1 in approximately every 23 million people. If we continued typing people until we reach 23 million, we would not expect to find someone else that matched for those four probes until after we had reached or exceeded 23 million people.[45]

Similar comments were made in other cases involving DNA,[46] blood,[47] and hair analyses.[48]

To estimate the number of people who would need to be tested before we might expect to find a match on a trait common to one in X people, we must compute the smallest N such that $(1 - 1/X)^N < .50$.[49] Thus, for F(Traits) = one in 100, we would expect to find a match after testing 69 people.[50] If 100 people were tested, the probability that at least one would match is about 63%.[51] Because the N that satisfies the equation will always be smaller than the denominator of F(Traits), the numerical conversion error exaggerates the

45. Ross v. State, No. B14-90-00659 (Tex. App. Feb. 13, 1992) (transcript at 129-30).

46. *See* transcripts from Jones v. State, 569 So. 2d 1234 (Fla. 1990), and Perry v. State, 586 So. 2d 242 (Ala. 1991).

47. State v. Erickson, 363 N.W.2d 859, 861 (Minn. App. 1985).

48. United States v. Jefferson, 17 M.J. 728, 734 (N.M.C.M.R. 1983).

49. The .50 level is used because an event whose probability is greater than 50% is "expected." For computational simplicity, we may solve for N where $N = \ln(.50)/\ln(1 - 1/x)$.

50. P(Another Match) = $1 - .99^{69} = 50.02\%$.

51. P(Another Match) = $1 - .99^{100} = 63.4\%$.

number of people who would need to be tested before a match may be expected. This, in turn, exaggerates the probative strength of the DNA match.[52]

V. WHY DO THESE ERRORS OCCUR?

Having identified errors that can and do occur in connection with DNA evidence, it is important to consider why these errors persist. Admittedly, the conversational context of the attorney-expert exchanges makes it difficult for the expert to catch and correct all probabilistic distortions. But given that all the errors appear to exaggerate the probative strength of DNA and other identification matches, it is reasonable to consider various motivational theories for their appearance. Could it be that the probative strength of DNA evidence is deliberately exaggerated by forensic experts interested in puffing the utility of their science, or prosecutors determined to win their cases? A review of trial transcripts suggests otherwise. DNA experts generally begin the statistical portion of their testimony with statements about estimated population frequencies and comparisons with a "random man." Broader, misguided statements about source and guilt probabilities typically emerge only after the experts redescribe or expand upon their initial testimony in response to attorneys' questions.

Motivational explanations are also weakened by evidence that the errors are routinely committed even by those who would seem to be motivated *not* to commit them. For example, the source probability errors in *Pierce* and *Thomas*, the ultimate issue errors in *Bethune* and *Womack*, and the P(Another Match) error in *Spencer* were all committed by defense attorneys who surely had no incentive to exaggerate the strength of the evidence against their clients.

A better explanation for the plethora of errors is the simplest one: ignorance. Few jurists are trained in probability theory, and most DNA experts who testify at trial know a good deal more about DNA laboratory procedures than the subtleties of probabilistic inference. Indeed, there is a great deal of evidence that people have trouble differentiating probabilistic information from the probabilistic hypotheses that the information informs; such confusion is consistent with the commission of the errors identified here and supports the ignorance hypothesis.[53] Finally, the well-known fact that everyone (save identical twins) has a unique DNA code may contribute to DNA evidence exaggerations; people may confuse that which is biologically inevitable with that which our technology is capable of revealing.

---

52. Numerical conversion errors will not be orders of magnitude greater than the correct answer. Whether they are sufficient to impact factfinders' judgments is an empirical matter.

53. *See* David M. Eddy, *Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities, in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman et al. eds., 1982); David H. Kaye, & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?* 154 J. ROYAL STAT. SOC'Y (A) 75 (1991); Willem A. Wagenaar, *The Proper Seat: A Bayesian Discussion of the Position of Expert Witnesses*, 12 LAW & HUM. BEHAV. 499 (1988).

VI. THE ROLES OF THE SCIENTIST

Clearly, there is profound confusion about the meaning and limits of the probabilities that accompany DNA matches and other identification evidence. Because intuitive judgment and legal policy do not provide accurate and consistent guidelines, the scientific community bears a special responsibility for educating the courts.

A. Scientists as Teachers

The law school curriculum provides a natural starting point. Scientists should work with the law school establishment to develop courses in scientific methodology, statistics, logic and probability theory. Judges' colleges should offer mini-courses in these areas as well. Jurors should be exposed to materials and testimony from court-appointed statisticians, logicians, or probability theorists in cases that involve important and complex quantitative evidence. Neutral advisory boards and clearinghouses specializing in quantitative testimony and materials should be established.[54]

Skeptics should take note of two points. First, it is not difficult to provide a legal context for lessons on probability and statistics. Case law is replete with wise and not-so-wise commentary on the application of quantitative methods and reasoning. Judges and lawyers who are exposed to both types of commentaries would be in the best position to identify fallacious reasoning when it comes their way. Second, evidence from the social science literature indicates that people may be trained to make quantitative inferences rather quickly and successfully. In one study, people who were exposed to a thirty minute lecture on reasoning performed better on quantitative inference tasks in a variety of domains than those who were not exposed to the lecture.[55] Similarly, another study showed that mock jurors who were exposed to a brief lecture on Bayes's theorem made better judgments than jurors who were not exposed to the lecture.[56] Though not definitive, these studies suggest that one need not have a Ph.D. in statistics or decision theory to avoid falling prey to the probabilistic errors identified here.

B. Scientists as Experts

Forensic scientists must be aware of the meaning and limits of their testimony. Specifically, they must not draw overtly probabilistic inferences

54. *See* THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS 13 (Stephen E. Fienberg ed., 1989).

55. Richard E. Nisbett et al., *Teaching Reasoning*, 238 SCIENCE 625 (1987).

56. Brian Smith et al., Bayesian Presentations and Juror Use of Probabilistic Evidence (Mar. 1992) (unpublished paper presented at American Psychology-Law Society Biennial Meeting, San Diego).

from their results about whether a suspect is the source of a trace, was present at a crime scene, or committed a crime. Although matching evidence is usually probative with respect to these issues, such inferences cannot be made solely on the basis of laboratory tests.

Not only should forensic scientists avoid misstatements, but they should remain sensitive to the ways in which their conclusions will be interpreted. Speculation about probabilities associated with analyses not yet conducted is potentially confusing and prejudicial.[57] Likewise, statements such as "[t]here is no evidence that the semen of the vaginal swab originates from anyone other than Wayne Amundson,"[58] are highly misleading. Such statements tie the diagnosticity of trace evidence to the number of people tested by the laboratory; the fewer people tested, the greater the chance that a second match will not be found.

Sensitivity to factfinder interpretation of DNA testimony also requires forensic scientists to point out the inferential limitations of their tests. They should acknowledge that coincidental matches are possible. They should admit the complexities associated with the construction of reference populations. They should stop insisting that false positive errors are impossible. They should be prepared to discuss proficiency test procedures and results for the industry and their own laboratories.

One strategy to identify error rates would be to require the forensic scientists to state the *highest* error rate that is consistent with laboratory performance in realistic, blind, external proficiency tests to date.[59] Thus, if a laboratory commits zero false positive errors out of, say, 100 match reports, jurors would be informed that such performance is consistent with a true error rate of 3% or less (based on a 95% confidence interval).[60] Although conservative, this policy creates an incentive for laboratories to participate in proficiency tests that could reduce the reported upper bound.[61] In the absence of substantial proficiency test data from a particular laboratory, the proficiency test data could be combined across laboratories to estimate the false positive error rate

---

57. One recent newspaper article reports an FBI scientist testified that an F(Traits) = 1/466 would rise to one in 3 billion if additional tests matched. But until a match is found, such statements are irrelevant. Jerry Shnay, *DNA Tests Tie Moore to Woman's Slaying*, CHI. TRIB. June 5, 1992, at 4.

58. Jonathan Gaw, *Wis. Man Extradited in Prostitute Attacks*, L.A. TIMES, July 1, 1992, at B4 (quoting from case report of the Serological Research Institute of Richmond, California).

59. Michael J. Saks & Jonathan J. Koehler, *What DNA "Fingerprinting" Can Teach the Law About the Rest of Forensic Science*, 13 CARDOZO L. REV. 361, 369 (1991).

60. For the general case in which p is the true (but unknown) positive hit rate and r is the number of positive hits in proficiency testing out of n reported matches, the upperbound false positive error rate for a 95% confidence coefficient is found by subtracting from one the largest p for which $[n!/r!(n-r)!]p^r(1-p)^{n-r} \leq 0.05$. Where $n = r = 100$, $p \geq 0.97$. Hence, the error rate is 3% or less.

61. For a laboratory that never makes false positive errors in proficiency tests, the upperbound false positive error rate estimate drops from approximately 5% for n = 50, to 1% for n = 300.

of any particular laboratory. Analysts who believe their own error rate is smaller than average would have the burden of justifying their claim.

C. Scientists as Researchers

Scientists also have a duty to improve the use of forensic statistics at trial by conducting research into a number of key issues. The subpopulation issue already is the subject of active research, and there is growing awareness that more and better proficiency tests need to identify forensic science error rates.[62]

Until now, little attention has been paid to the psychology of match declarations and the impact of reported matches on decision makers. The standards for declaring a match are variable, and a laboratory analyst's expectations or goals may influence his or her match determination. Some forensic scientists have testified that the discovery of a match at one locus helps determine whether a close call at another locus will be declared a match.[63] Presumably, a match declaration at the first locus increases the analyst's confidence that the suspect is the source, and the analyst's expectation of finding a match at other loci increase as well. This expectation might then be used to declare matches in ambiguous situations that otherwise would be declared nonmatches. If so, the independence assumption used to compute F(Traits) across a series of loci would be invalid. In some cases, forensic scientists have testified that it would not be possible for match declarations to be affected by an analyst's expectations or knowledge about the strength of the additional evidence in the case.[64] Regardless of where the truth lies, forensic scientists' beliefs about influences on their judgment are a poor substitute for empirical investigation. We must investigate what impact, if any, analysts' hopes, expectations or perceptions about the strength of a case against a suspect have on their judgments about the genetic evidence.

As for the impact of match declarations on factfinders, it is widely assumed that reported DNA matches are decisive—or at least extremely persuasive. But the impact of DNA evidence on factfinders is an empirical question whose answer will likely depend on the context and way in which the evidence is presented. Recent studies showed that mock jurors who were provided with source probability statements about forensic science evidence reported higher probabilities of guilt and were more likely to convict than were mock jurors who were provided with random match statements.[65] Similarly, different fram-

62. NATIONAL RESEARCH COUNCIL, COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992); Proposed Crime Control Act of 1993 § 1364, 103d Cong.

63. Lander, *supra* note 4, at 503 (quoting Michael Baird).

64. United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991).

65. Koehler, *supra* note 22; Jonathan J. Koehler et al., The Source Probability Error in the Presentation of DNA Evidence (unpublished manuscript on file with author). This effect was most pronounced for F(Traits) = 1/1,000. The effect diminished significantly for F(Traits) = 1/1,000,000,000.

ings of blood evidence in combination with arguably fallacious statements about this evidence was reported to have a large impact on mock jurors' probability of guilt estimates.[66]

More precise specification of the conditions under which misrepresentations of forensic science evidence will have a significant impact on factfinders is needed. For example, will the use of visual analogies or some related cognitive aid to convey the diagnostic significance of F(Traits) increase the impact of a reported match on the factfinder? Will informing jurors about laboratory error rates and protocols affect their judgments about the diagnosticity of the forensic evidence? Will these diagnosticity judgments affect jurors' evaluations of the strength of nonforensic evidence? Studies that examine these issues can provide the courts with a much-needed empirical basis for gauging the practical significance of the errors and exaggerations that occur when identification evidence is presented.

66. Thompson & Schumann, *supra* note 36.