1991

Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes

Richard O. Lempert  
*University of Michigan Law School, rlempert@umich.edu*

Available at: [https://repository.law.umich.edu/articles/2383](https://repository.law.umich.edu/articles/2383)

Follow this and additional works at: [https://repository.law.umich.edu/articles](https://repository.law.umich.edu/articles)

Part of the Evidence Commons, and the Forensic Science and Technology Commons

**Recommended Citation**


This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.
SOME CAVEATS CONCERNING DNA AS CRIMINAL IDENTIFICATION EVIDENCE: WITH THANKS TO THE REVEREND BAYES

Richard Lempert*

INTRODUCTION

The conference panel at which this paper was originally presented was structured along the lines of a debate. The three speakers who were supposed to advocate the use of DNA evidence were labeled, as is customary, Proponents. But those who were supposed to take the negative side were not called Opponents. Rather they were labeled Caveators. I do not know who is responsible for this label, but I think it gets things exactly right. To my mind anyone considering DNA as criminal identification evidence should be a Caveator. The promise and utility of DNA analysis in identifying the perpetrators of such serious crimes as rape and homicide must be acknowledged, but one must also be aware of the limits of the DNA identification process as it now exists and the ways in which these limits affect what experts can reliably tell judges and jurors. In particular, I shall argue below that current practices may lead to misleading claims for reasons that to date have not been fully appreciated by the forensic science community. In making some of these arguments I shall use the Bayesian perspective which Finkelstein and Fairley long ago posited as a paradigm for thinking about identification evidence.

Although the forensic use of DNA to identify criminals is in this country less than five years old, the law review literature is already replete with articles discussing the technology. Many of these arti-

---

* Francis A. Allen Professor of Law, University of Michigan Law School. B.A., 1964, Oberlin College; J.D., 1968, Ph.D., 1972, University of Michigan. Work on this paper was supported by the Cook funds of the University of Michigan Law School. I especially thank Ian Evett and William Thompson who commented helpfully on the initial version of this paper and who provided me with material that was of great aid as I revised. This version of the paper is, however, very different from the version they saw, and there is no reason to expect that they would agree with everything I have written. I am also grateful to Sam Gross and Larry Kramer for their helpful comments on this manuscript.

1 I assume Peter Tillers who organized the conference is responsible for the label, but I never went so far as to ask him.


3 See, e.g., Thompson & Ford, DNA Typing: Acceptance and Weight of the New Genetic Identification Tests, 75 Va. L. Rev. 45 (1989); Note, The Dark Side of DNA Profiling: Unrelia-
cles include detailed descriptions of how identification by DNA matching proceeds. Thus I see no reason to be similarly detailed here.\(^4\) I should note, however, that while some of my comments apply Scientific Evidence Meets the Criminal Defendant, 42 STAN. L. REV. 465 (1990) (authored by J.C. Hoeffel) [hereinafter, Note, DNA Profiling]; Note, DNA Typing: A Rush to Judgment, 24 GA. L. REV. 669 (1990) [hereinafter Note, DNA Typing].

\(^4\) The custom in law review writing is to assume no knowledge on the part of the reader and to build every argument from the ground up, even if that ground has been plowed many times before. It is a custom that has wasted numbers of trees. Contrary to custom, I assume that most of my readers are already familiar with the basic mechanisms and criticisms of DNA identification, for my purpose is not to educate readers in an unfamiliar technology, but to push an ongoing discussion further. Those who are not well acquainted with DNA identification in its forensic context may wish to peruse other articles examining DNA identification evidence before turning to this one. One article which gives an especially good overview of the technology and raises many of the most basic criticisms that have been raised against these procedures is Thompson & Ford, supra note 3.

For those readers who have not yet looked at the literature on DNA identification but nonetheless wish to read on, the following, without pretending to be an adequate explanation of the technology of DNA analysis, is all that one must know to understand the arguments in this paper.

A typical DNA identification proceeds by comparing DNA left at the crime scene (extracted from the evidence DNA sample—most commonly semen) with DNA known to belong to the suspect (extracted from the suspect DNA sample—most commonly blood taken from a suspect). The DNA taken from the two samples is cut into fragments at particular places (loci) by restriction enzymes and these restriction fragments, called RFLPs (restriction fragment length polymorphisms) or alleles (one of two or more forms of a gene at a gene locus) are compared by measuring their length. This measurement is done by placing the DNA in a gel and running an electric current through the gel. This causes the negatively charged DNA to migrate through the gel, but the larger the allele the slower the migration process. Thus when the current is turned off, shorter strands of DNA will have moved farther than longer ones. These differences in movement can be observed by comparing the locations of what may be described as fuzzy lines (or, sometimes more accurately, elongated blobs) called bands on an autoradiogram, or autorad for short. (For the purposes of this paper, one need not know how the bands were made to appear on the autorad.)

If the alleles from the evidence and suspect samples are genetically the same, they will be the same length, which is to say they will have moved the same distance, plus or minus measurement error. This identity will ordinarily be revealed on the autorad because the bands in a column (lane) that represents evidence DNA will be at the same location (that is, the same distance from the bottom of the autorad) as the bands in a column that represents the suspect DNA. If the alleles are different, they are likely to be of different lengths, but it may be impossible to distinguish two different alleles because they may be so close in size that the observed differences in length could plausibly be attributed to the errors that might occur in twice measuring identical alleles. In other words, different size alleles may appear to be located at virtually the same point (give or take measurement error) on the autorad.

If the alleles from the evidence and the suspect sample are the same length one must know to what degree this incriminates the suspect. This depends on how many people are likely to have apparently identical alleles at the places examined. If 90% of all people have identical DNA at a particular locus, the fact that a suspect's DNA matches a rapist's DNA does little to inculpate the suspect. If on the other hand, only one percent of all people have DNA at a particular locus, the fact that a suspect's DNA matches the evidence DNA at that locus carries substantial incriminatory weight.

The judgment about how rare different-sized alleles are is made by looking at the frequency of different alleles in a more general population. Laboratories that measure different
ply more broadly, I have in mind only the current dominant technology, RFLP analysis using single locus probes with Southern blotting.

I. PROPER STATISTICS

DNA identification evidence is ordinarily presented to the fact finder as the fact that a sample of a defendant's DNA matches or does not match an evidence sample and, if it matches, the probability that an individual randomly selected from a population would have DNA matching the evidence sample. In the United States, experts typically give this probability in frequentist terms; for example, "there is one chance in fifty thousand that a randomly selected Caucasian

alleles will each have done population studies in which the length of the different alleles that may be found at the loci of interest will have been measured for a sample of at least several hundred individuals. Indeed, typically several such population studies will have been done, usually one for Caucasians, one for Blacks, and one for Hispanics. The suspect's race will determine which population base is used to indicate allele frequencies. In cases where the suspect DNA matches the evidence DNA, the proportion of the population sample with the same alleles will determine the incriminatory weight of the evidence.

Because different alleles may be inherited from each parent, there will often be two different alleles at a loci, and since a typical DNA identification may examine alleles at four loci there may be as many as eight opportunities for the evidence and suspect DNA to match. If there are no test or measurement problems, the suspect cannot be the source of the evidence DNA unless all alleles tested match up. Thus if an allele or alleles at a single loci do not match up, and if the failure cannot be attributed to some condition of the test or measurement, the suspect cannot be the source of the DNA found at the crime scene. If, on the other hand, there is a perfect match, this fact may or may not have substantial incriminatory value. The degree of incriminatory value depends on the proportion of people in the appropriate population who have the same configuration of alleles (i.e., a joint distribution of alleles) across the loci that have been examined. However, given the size of available population samples, it is likely that when four loci have been studied no one in the population sample will have exactly the same allele configuration as that found in the suspect and evidence samples. Thus, the probability of the joint distribution must be estimated from the frequency of individual alleles in the population sample. Typically this is done by assuming that each identified allele exists independently of the existence of each of the other identified alleles. This means that the probability of a particular distribution of alleles may be calculated by multiplying together the probabilities that each individual allele will be found. The assumption of allelic independence is, however, not justified in a world where people do not mate at random, (see sources cited supra note 3; infra note 9) although there is considerable dispute over what the failure of this assumption implies (see sources cited infra note 76). In cases where suspect DNA matches evidence DNA, prosecution experts who assume allelic independence and multiply individual probabilities have estimated likelihoods of one in fifty billion and less that a randomly selected person would have DNA matching the DNA found at a crime scene and upon analysis of the suspect's blood.

5 The fact finder may be a judge or a jury, and even when the jury is the fact finder the judge too must find facts, for he/she must determine whether the jury's fact finding, in the case of a guilty verdict or any verdict in a civil case, is reasonable enough to be allowed to stand. Nevertheless, I shall assume for purposes of exposition that the fact finder is a jury.

6 I will assume it is the defendant's DNA that is being assessed. The DNA sample might, of course, come from someone who is not a suspect, the boyfriend of a rape victim, for example.
male would have the same DNA profile as that found in both the evidence sample and the sample taken from the defendant.” This frequentist probability does not, however, directly answer the question that confronts the jury, which is not how unusual is the DNA profile in question, but how likely is it that the evidence sample is that of the defendant? The answer to this question will turn, even in the frequentist world, on the size of the population of potential suspects. Unfortunately, the careless presentation of evidence, either by an expert witness or by a prosecutor summarizing the expert’s testimony for the jury, may make it look as if the question of the rareness of the evidence DNA profile and the probability that the defendant’s matching DNA is the source of the evidence profile are identical. To continue with the above example, the jury may be led to believe that the import of the DNA evidence is that there is only one chance in fifty thousand that the evidence DNA came from someone other than the defendant.7 This is known as the “prosecutor’s fallacy.” And even if this


For an example in a case involving a DNA identification, consider the following colloquy between judge and expert in a Frye hearing in the case of People v. Johnson, No. A 998 149 (Los Angeles Super. Ct. Aug. 22, 1990). The expert is a molecular biologist who worked for Cellmark Diagnostics, one of the private laboratories that has been a leader in moving DNA identifications from the laboratory to the courtroom. The issue in the case is somewhat atypical. The question concerned the identity of a missing man, whom the police believed had been murdered by the defendant. To establish identity the state sought to show through DNA testing that blood flakes found in the defendant’s van belonged to the missing father of someone whom I will call X. X’s father was believed to have been the defendant’s accomplice in a robbery.

Prosecutor: As a result of the testing you’ve done in this case, what is your answer? Is it that the testing shows that the blood flakes are consistent with being the biological father of the son, that is [X], or that in fact the blood flakes as a result of your testing and in your opinion do represent the biological father of [X]?

Witness: In my professional opinion, the biological relationship between [X] and the blood flakes is a true and accurate one. The data I say is consistent; however, that is what—in my professional opinion there, is a biological relationship there.

The Court: Does that mean it’s consistent with his being a father or it’s 100 percent certain that he is the father?

Witness: In my opinion that he is the father.

The Court: That the flake is the father of [X]?

Witness: Yes.

The Court: Because you compared [X’s] blood with [X’s mother’s] blood and the flake’s blood?

Witness: Yes.

The Court: Okay. I apologize. I misunderstood your answer.

Record at 309-10, Johnson.

In fact, the judge understood the limits of what the witness could say on the basis of the
misinterpretation is not offered to the jury, the jury on its own may make this mistaken equation.

Proper weighing of evidence by the jury requires a proper understanding of what the evidence imports. To achieve this a court must carefully monitor the way in which statistics are presented and interpreted by both expert witnesses and counsel. Instructions on the meaning of the evidence and a defense counsel adept at explaining how the evidence relates to the question the jury must answer (i.e., "how likely is it that the defendant is the source of the evidence DNA?") may also be necessary if a jury is to properly weigh the evidence it is given.8

If the defendant's DNA matches the evidence DNA, the question of how likely it is that the defendant is the source of the evidence DNA turns on the composition of the population of suspects. Thus it is clear that assumptions about the distribution and independence of alleles in the reference population used to evaluate the probability of a DNA match is crucial. It is now well recognized that due to population substructure, standard population data bases may not adequately represent the relative frequencies of polymorphic alleles in specific subpopulations.9 Thus data bases consisting of American Blacks might not adequately assess the likelihood that the evidence DNA profile would characterize a randomly selected West Indian Black when the defendant whose DNA matches the evidence DNA is a West Indian. Similar arguments may be made about extending analyses from Hispanic data bases to, for example, Cubans or generalized Caucasian data bases to Sicilians.10

---

8 Instructions and explanations might not be crucial where the probability that a random individual possesses the DNA profile in question is very low, such as one in ten billion. But while probabilities of this magnitude have been offered in litigation, current statistical data bases and the current status of the assumptions that alleles are independently and randomly distributed in data base populations are insufficient to support estimates of such magnitudes given the number of alleles ordinarily evaluated.


10 Whether subpopulation data matters depends on the population substructure of the data base and suspect populations. See infra notes 20-21 and the text accompanying notes 16-21.
Juries are not, however, in a good position to evaluate the problems posed by such subpopulations in their weighing of evidence. While a jury can appreciate the fact that it may be improper to apply statistics based on population data to members of a distinct subpopulation, there is little that a jury can do rationally to weigh the implications of this fact because with only population data the jurors have no basis for estimating the probative value of a match. Rather the problem must be resolved at the level of the standards that are set for qualifying experts, for population data bases, and for appropriate conservatism in making statistical estimates.

A. The Problem of Micro Populations

While considerable attention has been paid to problems posed by subpopulations that may not be homologous with the population of which they are a part, virtually no attention has been paid to the problem posed by "micro populations"; that is, the group of brothers, sisters, cousins, aunts, uncles, parents, and children who reside in the same general area as the defendant and who might be possible sources of the evidence DNA. For example, in a tour of one DNA laboratory I saw one autorad in which there was a perfect match except for two clearly disparate alleles, which resulted in an absolute exclusion. The DNA analyst reported that he sent a message back to the prosecutor who had submitted the sample that the defendant should be freed and his brother arrested.

Presumably if a brother had been arrested and his DNA matched the evidence DNA perfectly, the expert would have been willing to testify to some high degree of probability that the DNA could not have come from some randomly selected individual, and the jury would have been invited to infer that this meant that there was a high likelihood that the new defendant was the source of the evidence DNA. But in recommending that a brother be arrested, the expert never drew the prosecutor's attention to a crucial question; namely, how many brothers did the defendant have? If the defendant had ten brothers, the likelihood might be quite high that more than one would match the evidence DNA perfectly. Cousins, children, parents and other close relatives would raise still further the probability that there was at least one other person in the vicinity who was not excluded as a suspect by the DNA evidence, meaning that the probability that someone other than the defendant was the source of the evidence

The estimated probability that a random individual might be responsible for the evidence DNA is likely to be higher if the data base used to make this evaluation is composed of individuals who are ethnically of the same subpopulation as the defendant.
DNA might have been as high as one in two or one in three or some other similar number. If the matter is called to their attention, juries might be able to understand the import of the fact that there are untested close relatives who are potential suspects, but they cannot properly weigh the implications of this possibility unless the scientific community pays some attention to the question of what the presence of micro populations of close relatives means when interpreting DNA evidence for forensic purposes.

Ian Evett, who is to my knowledge the only forensic scientist to have formally addressed the matter,\(^1\) acknowledges the potential seriousness of the problem. He posits, by way of example, a situation where finding a match between evidence and suspect DNA leads to a two hundredfold increase in the estimated odds of a defendant's guilt. This increase (which may be thought of as the evidentiary weight of the match) diminishes to less than two percent of its original value if the defendant's brother is a plausible suspect.\(^2\) Evett, with typical British understatement, notes that “Non-scientists [read “jurors”] might not be aware that such changes [in the weight of the evidence] are possible.”\(^3\) One might add that scientists, especially forensic scientists testifying to the weight that should be given DNA matches, seem not to be aware of this either; at least I have encountered no reported case in this country where a prosecution's expert discounted the evidentiary weight of a DNA match because of the possibility that a relative might have left the sample.

While this micro population issue most obviously arises when the defendant claims that the perpetrator was his brother or some other relative, it in fact exists in almost every case. For it is the state's responsibility to ensure that its expert's estimate of the weight to be placed on a match is based on appropriate population statistics. This is part of what it means to have a presumption of innocence and to place the burden of proof on the state.\(^4\) In other words, before introducing any sort of statistical identification evidence, it is the prosecutor's obligation to identify the appropriate population base for the

---


\(^2\) I. Evett, supra note 11, at 5-6.

\(^3\) Id. at 6.

\(^4\) The burden of proof is on the state in two senses. First the state bears the risk of non-persuasion in that if the jury is not persuaded of the defendant’s guilt beyond a reasonable doubt, it should acquit. Second, the state has the production burden or the burden of presenting enough evidence to allow a jury to find guilt beyond a reasonable doubt. In meeting this burden the state may not present testimony that it knows to be false and must base any scientific evidence it offers on a scientifically appropriate foundation.
Thus it is the prosecutor's obligation from the outset to specify in a scientifically justifiable way what that population is. If a brother or other relative is a member of that population, an expert's opinion of the probability that the DNA came from someone other than the defendant should take account of this fact.

The population that should be used to interpret the evidentiary implications of a DNA match is the population of potential suspects, which I will call the "suspect population"; that is, the group of people who might have committed the offense if the defendant is not guilty. Members of the suspect population will typically be those people who are in a position to have committed the offense (predominantly people who live close enough to the crime scene to have been there at the time of the crime) as limited by other known facts that might screen the population potentially subject to DNA testing. If, for example, a killer had only one leg, only people who might have been present at the crime scene and are missing a leg would be members of the suspect population, for before the state went to the expense of testing a person's DNA they would check to be sure he had only one leg. Or, to be less fanciful, if a raped woman did not know whether her assailant were white or black because she was assaulted from behind and then blindfolded, the suspect population would consist of the population of people, white and black, who could have been present at the time and place of the assault. The appropriate data base for evaluating the weight to be placed on a match, would be a data base which mixed white and black cases in proportion to their presence in the suspect population. Conversely, if the victim had identified her rapist as white, the appropriate data base for evaluating a match would be a data base composed entirely of men who appeared to be white, since the police would not arrest and test a man who appeared to be black in these circumstances.

DNA experts never use data drawn from the suspect population.\(^\text{15}\) Some other people may also be members of this population. For example, a tourist who was in the area of the crime at the time is a member of the population of suspects though he may not be in the area at the time the crime is investigated and no one may appreciate that he is a suspect. For most purposes such people may be ignored in deriving appropriate population statistics. In some cases, such as mob hit men, who may frequently be from out of town, or high level drug distributors, people who reside outside of the area of the crime cannot be ignored and, indeed, are not ignored in police investigations.\(^\text{16}\)

\(^{15}\) See, e.g., People v. Collins, 60 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

\(^{16}\) See, e.g., People v. Collins, 60 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).
to evaluate the implications of a match. Rather they use data from the same convenience samples as a base against which to evaluate all the DNA tests they conduct. This is justified by the claim that the DNA alleles evaluated are "unexpressed" and so should be distributed in one group (the group used to construct the data base) with a frequency that is no different, after accounting for sampling error, than the frequency with which it is found in another group (the suspect population). This assumption is typically qualified in one way. The data base used to determine the probability of a random match is composed of people of the same gross ethnicity (White, Black, Hispanic) as the defendant even though this limitation is appropriate only when other available evidence limits the suspect population to people of the defendant’s ethnicity.

With this understanding of the suspect population, we can better

---

18 "Well," in the words of Captain Corcoran to his crew, "hardly ever." In a well-known English case the police sought to test and largely succeeded in testing every male between the ages of 13 and 30 living in an area where two rape murders had occurred. The case eventually broke when one man confessed that he had provided a blood sample for another. The other's blood was then tested and his DNA matched that of the rapist. See the article Regina v. Pitchfork, London Times, Jan. 23, 1988, at 1, col. 1.

19 Unexpressed DNA alleles have no known manifestations at any level above the DNA and so could not affect survival rates or be used as a cue for mating.

20 To recognize this is to recognize population substructure as a factor which influences the distribution of the alleles tested. However, crudely controlling for ethnicity along the most salient socially defined lines fails to come to grips with the issue of population substructure that is implicitly recognized. The broad social categories in which ethnicity is socially defined do not define groups that are genetically homogenous. The distribution of alleles in a group of people with Finnish ancestry may be quite different from the distribution of alleles found in a group with a Lithuanian Jewish heritage, although we might label both Finns and Lithuanian Jews Caucasians. Indeed, the allele distribution within one white ethnic group may be closer to the allele distribution among all American blacks than it is to the allele distribution within some other white ethnic group. Moreover, there may even be structure to what we think of as subpopulations. For example, Italians might be considered one of the subpopulations or ethnic groups that make up the Caucasian population. But within contemporary Italy, population substructure exists which follows linguistic and to a lesser extent geographic lines. Barbujani & Sokal, supra note 9. Thus if a suspect population consisted largely of Sicilians, the estimated probability that a randomly chosen member of that population would have a certain configuration of alleles might be off by orders of magnitude if the estimated allele frequencies were based on an Italian rather than a Sicilian reference group.

21 Where the suspect population is not limited to people of one race, limiting the data base sample to individuals of the defendant’s gross ethnicity is conservative in the sense that it will understated the weight of the evidence if the suspect’s alleles are more commonly found in a group that shares his ethnicity than in a group that has a different heritage. I do not, however, think that this limitation exists because forensic experts seek to be conservative in the probabilities they present (though they sometimes do). Rather I think this is done both because little thought is given to the proper definition of the suspect population and because it is felt that juries will think it appropriate to compare individuals of one ethnicity with their "own kind." If the latter is a reason, socially constructed ethnic categories serve well even if they are incoherent from the point of view of population biology and genetics. For a discussion and an example of how to estimate allele frequencies for a racially mixed subpopulation, see Walsh &
understand the problems posed by brothers, cousins, fathers, and other relatives. More often than not some relatives will be members of the suspect population. Not only is a suspect likely to have relatives living in his vicinity, but the relatives are likely to be similar on initial screening dimensions such as race. Thus if the state properly defines the suspect population, however unlikely it is that a random individual could have left DNA that matches both the suspect’s and the evidence DNA, the probability that a member of the suspect population could have left matching DNA will often be much higher because relatives tend both to belong to the suspect population and to share alleles with the suspect.

The question is what to do about this. One solution is to live with the fact that evidence of a DNA match is not nearly so probative as people have thought because suspect populations are not random agglomerations with respect to the likelihood of sharing the alleles compared in DNA analysis. Totally apart from issues of population substructure, the likely presence of a defendant’s relatives in the suspect population means that there is a good possibility that another potential suspect has DNA that matches the evidence DNA. Living with this situation means telling juries in many instances that a DNA match has only a moderate rather than a strong incriminatory implication. I expect, however, that this solution will strike many people as inadequate. Their intuition will be that suggestions that “a brother did it” will more often than not be a red herring which will improperly undercut the potential of DNA analysis to provide relatively unambiguous evidence of identity. It is an intuition I share.

A second solution is to ignore the presence of relatives in the suspect population unless the defendant makes a plausible case that a relative committed the crime. This solution, however, runs counter to the general principle that the burden of proof is on the state and that it is the state’s burden to provide scientifically reliable estimates of the likelihoods involved in DNA analysis regardless of the claims the defendant makes. Moreover, unless the defendant were provided with the funds needed to conduct an in-depth criminal identification, he will ordinarily be in no better position than the state to make a plausible case that a relative of his committed the crime of which he is accused.

A third solution is for the state with the cooperation of the defendant\(^{22}\) to identify those relatives who belong in the suspect popul-

---

\(^{22}\) If the defendant does not cooperate in identifying relatives, I think the state should be

tion and for the state by examining their DNA to individually exclude them as suspects. This is, in principle, a sound strategy, but in many cases it will be expensive and inconvenient, with the further drawback that the process of investigating relatives in connection with a serious crime is likely to cast undue suspicion on numbers of innocent people. Moreover, the best and most private way of excluding relatives as suspects, which is to analyze their DNA, will depend on the relatives' cooperation since the mere fact of being related to the prime suspect will not generate probable cause for such an intrusive search. Guilty relatives as well as those who simply want to undercut the state's case against the defendant are unlikely to consent to DNA tests or even to cooperate in explaining their whereabouts at the time of the crime in question.

A fourth solution is to ask how likely is it, given the non-DNA evidence against the defendant and a DNA match, that a relative in fact left the DNA. If it is unlikely, the practical solution is to ignore the problems posed by relatives. In other words, if the non-DNA evidence in the case is strong (e.g., the defendant was caught by the police running from the scene of a rape, a footprint at the scene of the assault matched the shoes he was wearing, etc.), the presence of relatives in the suspect population can be safely ignored because it is unlikely that the DNA left at the scene came from anyone other than the defendant even if some relative's DNA matches the defendant's DNA and the evidence sample at the crucial loci. However, this solution means that DNA evidence will be reported as most weighty when it is least necessary; i.e., when the other evidence standing alone is likely

23 Even without probable cause individuals may be required to produce certain information, including personal information, in response to a subpoena issued in connection with the work of a grand jury. The leading case in this area is United States v. Dionisio, 410 U.S. 1 (1973), which held that the fourth amendment does not prevent a federal grand jury from securing a voice exemplar from an individual without a showing of probable cause. In reaching this decision the Court wrote, "The required disclosure of a person's voice is thus immeasurably further removed from the Fourth Amendment protection than was the intrusion into the body effected by the blood extraction in Schmerber." Id. at 14 (emphasis added). Even if DNA may be acquired through a less intrusive technique than the extraction of blood (such as the clipping of hair) the seizure of a DNA sample may still be different from acquiring a voice sample in that a person has a reasonable expectation that information about his genetic make up (unlike information about the sound of his voice or the color or texture of his hair) is private. Whether this argument, if accepted, should apply to VNTRs, which are unexpressed portions of the genome, is an issue I note here but shall not discuss.

24 Ignoring relatives will also mean that the statistic providing the probability that a random individual's DNA will match the evidence DNA is hypothetical, since we know the suspect population contains at least one nonrandom (with respect to DNA) individual—the
to support a conviction. Moreover, other evidence that we often think of as powerful may not have much probative value in this context. An eyewitness identification, for example, may suffer from the same weakness as a DNA match. Just as a relative of a suspect is more likely than a random individual to match that suspect's DNA, so is he more likely to be similar in appearance. Similarity means that an eyewitness presented with a lineup containing the innocent relative but not the guilty one has a good chance of picking the innocent relative out of the lineup. Thus an eyewitness identification would be insufficient reason to ignore possible relatives in establishing the probability of a DNA match.

Finally, technology may resolve the problem. If the number of sites tested (now usually three or four) is increased substantially, or if forensic scientists are eventually able to sequence DNA rather than merely measure approximately allele size, even the presence of close relatives (so long as they are not identical twins) in the suspect population should not prevent the conclusion that a DNA match provides overwhelming evidence that the suspect has been properly identified. But until technology advances, the most honest approach is to present the jury with the probability that a random individual left the evidence DNA and the probability that it was left by one of the group of defendant's relatives whom the state has not been able to exclude from the suspect population.

II. A BAYESIAN APPROACH TO THINKING ABOUT DNA

Additional problems in the interpretation and presentation of
DNA evidence arise, in part, because frequentist statistics are inherently unsuited to the forensic context. As I have pointed out above, the statistics usually presented to juries are not directly related to the question that the jury must resolve, and the relationship between the question the statistics can answer and the one the jury must answer may be unclear or prone to confuse. Putting aside for the moment the question of how statistical evidence should be presented to the jury, it is clear, as Finkelstein and Fairley argue, that the appropriate question, how much does a particular item of evidence increase the likelihood of the defendant's involvement in the crime charged?—is posed for the jury when statistical evidence of identification is presented in a Bayesian rather than a frequentist framework. Indeed, although evidence is presented in a frequentist fashion, jurors may attempt to evaluate it in a Bayesian light. And even if this is not true as a behavioral matter, it is how jurors should, as a normative matter, assess the probative value of evidence. Thus, Bayes' Theorem provides, at a minimum, a heuristic device which al-

\[\text{Evidence Strength} \quad \text{Likelihood Ratio In the Range of:}\]

<table>
<thead>
<tr>
<th>Weak</th>
<th>1 - 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair</td>
<td>33 - 100</td>
</tr>
<tr>
<td>Good</td>
<td>100 - 330</td>
</tr>
<tr>
<td>Strong</td>
<td>330 - 1000</td>
</tr>
<tr>
<td>Very Strong</td>
<td>1000+</td>
</tr>
</tbody>
</table>


28 I am not saying that frequentist statistics should not be used in a forensic context nor necessarily that the problems that inhere in them are severe. I wish simply to point to a tension inherent in their use; the fact that, as I have already noted, see text accompanying notes 6-7, frequentist statistics do not respond directly to the question that the fact finder must resolve.  

29 See supra text accompanying notes 5-8.  

30 I am referring to the frequentist statistics used by forensic scientists in the United States to describe for juries the degree to which a DNA match incriminates a defendant. The information need not be presented to juries in this fashion. In Britain, Ian Evett, a research scientist attached to the Home Office, has suggested the possibility of evaluating the probative value of DNA matches in a Bayesian perspective by constructing likelihood ratios. Under Evett’s proposed scheme a scientist would convey the evidentiary import of the evidence to the jury not in numbers, but in ordinary English. The conventions he has most recently suggested for translating likelihood ratios into words are:  

31 Finkelstein & Fairley, supra note 2.  


allows us to explore further issues relating to the proper weighing of DNA identification evidence.

For our purposes Bayes' Theorem may be written as:

\[ O(S|E) = \frac{P(E|S)}{P(E|\neg S)} \cdot O(S) \]

or, in English: the odds that the defendant is the source of the evidence sample, given the DNA evidence, equals the probability that the DNA evidence would have had the characteristics it had if the defendant were the source, divided by the probability that the DNA evidence would have had the characteristics it had if the defendant were not its source, multiplied by whatever odds one would have given prior to the receipt of the DNA evidence that the defendant was the source.

It will be immediately seen that in the case of a DNA exclusion \( P(E|S) \) equals zero, which means the ratio of probabilities, conventionally called the likelihood ratio, is zero, and the odds that the defendant was the source of the evidence DNA is zero regardless of how likely it appeared beforehand that the defendant was the source. Hence acquittals based on DNA evidence alone are justified.

This, however, assumes that one can be confident about an exclusion. Since an apparent exclusion might be attributable to various kinds of measurement error or DNA contamination, in practice \( P(E|S) \) will never be precisely zero and to conclude that the defendant is not the source of the evidence DNA and so deserves to be acquitted risks the danger of a mistaken exclusion, a risk that increases with the prior odds on \( S \). Thus the stronger the case against a defendant, the less the weight that should be given to a DNA exclusion.  

---

34 The FBI estimates that in slightly more than one third of the DNA identifications that they have been asked to do, the result of the test is to exclude the suspect. Personal Communication from Bruce Budowle, a Research Scientist associated with the FBI's Crime Lab to Richard Lempert and others (n.d.). This statistic should give one pause. DNA tests are presumably not ordered unless police have strong suspicion that a suspect committed a crime. If DNA exclusions are almost always accurate, it suggests that the police frequently arrest the wrong person. Conversely, police work may be better than it appears, and DNA testing may be conducive to false exclusions. In one Connecticut case the prosecutor apparently thought so, since a prosecution for rape proceeded despite a DNA exclusion and testimony by an FBI scientist in favor of the defendant. The jury convicted, choosing to believe the victim's eyewitness identification rather than the expert testimony. See Note, DNA Typing, supra note 3, at 690 n.104. It is possible that some DNA tests are done at the request of those who are only under mild suspicion, such as one spouse when the other has been murdered. If it is in these cases that DNA exclusions frequently occur, neither police work nor the accuracy of the test is undermined by the exclusion statistics. To date no one has looked at the kinds of cases that generate DNA exclusions and compared them to cases that yield DNA identifications.

Note also that the question of whether the defendant is the source of the evidence DNA is
It is also clear from the model that the implications of a DNA match for the probability the defendant is guilty ultimately depend on the prior odds that defendant is the source, which will often be thought to be the same as the prior odds that the defendant committed the allegedly criminal act. This has two important implications that can be missed when evidence is conceived of only in frequentist terms. First, the incriminating implications of a DNA match in the situation where the DNA is that of a suspect is very different and substantially greater than that of a similar match where the matched DNA belongs to an individual who but for the match would not be a suspect (for example, a person whose DNA profile happens to reside in a DNA data bank). In the first instance the weight of the evidence after the DNA identification will consist of already substantial prior odds multiplied by some high likelihood ratio. In the second case the prior odds on guilt will have been slight, and even after multiplication by a substantial likelihood ratio these odds may remain relatively small. Thus if there were no particular reason to believe that $X$ was any more likely to be guilty of a rape than any of one million other adult males who lived in the vicinity of the crime, the prior odds on $X$'s guilt would be $1:1,000,000$ and multiplication by the extremely incriminatory likelihood ratio of $100,000$ to $1$ would still leave odds on $X$'s guilt of $1:10$, odds more consistent with innocence than with guilt. On the other hand, if other evidence suggested that there was a fifty-fifty chance that $X$ was guilty, or odds of $1:1$, the same DNA test would leave posterior odds on guilt of $100,000:1$, far more than is needed for proof beyond a reasonable doubt.

Second, if jurors reason in anything like a Bayesian fashion, evidence already captured in the prior odds should not be replicated in the likelihood ratio, or else the final odds on guilt will be exaggerated by double counting. This can occur in the ordinary case because the equation as presented is somewhat misleading. Evidence does not present itself to the jury, but is presented by individuals as expert witnesses. The police in furnishing laboratories with DNA evidence often include information about the case against the suspect. Not the same as the question of whether the defendant is guilty of a crime. A defendant may leave no trace of DNA evidence in circumstances where because other evidence suggests the defendant's guilt, one would give high prior odds that the defendant is the source of the DNA found (e.g., a rape in which there is substantial evidence that the defendant is the rapist, but unknown to anyone the victim had intercourse with another person before the attack and the defendant rapist wore a condom). This is another reason why one should not regard a DNA exclusion as meaning that an acquittal is necessarily required. Conversely, the defendant may be the source of crime scene DNA although he is innocent of the crime. (E.g., a blood stain on the ground may have been left by someone who picked up a knife, accidentally cut himself, and then fled when after spotting a body he realized the knife was a murder weapon.)
expert's judgment regarding the existence and probative value of a DNA match is influenced by her knowledge of other incriminatory information, the jury's estimate of the incriminatory weight of the DNA evidence will be inappropriately high. If the incriminatory evidence that influences the expert is also given the jury, this evidence will be double counted, once explicitly for what it is worth and once unknowingly as part of the DNA evidence.\textsuperscript{35} If the other evidence is not given the jury; for example, if it were the fruits of an illegal search, the jury will nonetheless be considering it when it gives weight to the DNA evidence.

There is an irony here. An expert's judgment of a match is likely to be more accurate if the expert knows and is influenced by other evidence that the suspect is guilty,\textsuperscript{36} but an expert's judgment that is so influenced is likely to be misweighed by the jury. Moreover, an expert who is familiar with the evidence against the defendant and is influenced by this familiarity in calling or not calling a DNA match is presenting, under the guise of DNA testimony, a judgment that no expert is allowed to give; namely, a judgment on the probative value of scientific and unscientific evidence taken together. While the expert might claim that her knowledge of other incriminatory information would not influence her evaluation of the DNA evidence, so long as experts are human, such complete discounting is likely to be impossible wherever the DNA evidence is sufficiently ambiguous to allow conflicting interpretations.\textsuperscript{37} Thus there should be a general rule that

\textsuperscript{35} Even if the evidence the DNA expert is aware of and influenced by is not presented explicitly to the jury it may be double counted, for jurors may begin with prior probabilities of guilt that are based on the assumption that the defendant would not be in court if the police did not possess incriminating evidence not (or not yet) known to them.

\textsuperscript{36} For example, suppose an expert in a rape case is confronted with an apparent nonmatch but is not certain whether she can rule out the possibility of band shifting. If she knew that the suspect was found with a watch and other jewelry taken from the victim in the course of the rape and that the defendant had confessed to the crime when he was arrested, she would be more likely to perceive band shifting and a possible match than if she did not know this. See infra note 47. Moreover, testimony that the DNA matched or that the analysis was inconclusive is more likely to be correct than testimony that the DNA did not match. Yet if, but for her knowledge of the case, she would have testified that the DNA did not match, testimony that the DNA matched or that tests were inclusive, reports not a scientific judgment but the expert's opinion that viewing the DNA evidence in light of what the police told her, the suspect must be guilty. Moreover, the evidence that shaped the expert's testimony might be inadmissible at trial because the defendant's rights were violated in the process of acquiring it. While some might applaud allowing expert testimony to be shaped in this fashion as a way of getting around exclusionary rules they think silly, circumventing the rule in this way is constitutionally inappropriate.

those who do DNA testing should not know what other evidence in-criminates a suspect; indeed, until a judgment on the DNA has been reached, DNA analysts should not know the identity of the case in which they are involved or the side that has requested the test. This does not mean, however, that such experts should be deprived of all information pertaining to the crime. Some information, such as the number of assailants in a rape case, is needed by the expert to properly test the DNA and to interpret test results. Such information must be provided.

The danger of double counting is also an important reason why prosecutions based on matches with DNA found in DNA repositories pose special dangers of false convictions. If the DNA evidence is presented to the jury, it is likely to be assumed that it is cumulative of prior evidence that identified the defendant as a suspect rather than that it is the source of suspicion. Even if by the time of trial there is other evidence that the defendant is guilty, such as an eyewitness identification, it may be that the other evidence is "contaminated" by the DNA evidence, as it would be, for example, if an eyewitness learned that the person he/she was called upon to identify had been fingered by DNA evidence. 38 Thus where a person has become a suspect because evidence DNA matched his DNA in a data bank, evidence of that match should not be admitted unless, as with fingerprint evidence, the chances of the evidence DNA belonging to anyone else are minute. 39 If the jury is given the DNA evidence that generated suspicion, the evidence may well be misweighed and there is a non-trivial danger of a mistaken conviction if the other evidence in the case cannot by itself support a finding of guilt. 40

---

38 Even if the eyewitness did not know about the DNA match a contamination effect would be likely to exist since the defendant would probably be one of a small number of people presented to the witness as suspects, and considerable evidence suggests that eyewitnesses may be prone to identify suspects in suspect-absent situations. See e.g., E. Loftus, Eyewitness Testimony (1979); E. Loftus & K. Ketcham, Witness for the Defense (1991); Lindsay, Wells & O'Conner, Mock-Juror Belief of Accurate and Inaccurate Eyewitnesses, 13 Law & Hum. Behav. 333 (1989).

39 This will often be unlikely in the case of a data bank match. See infra note 90 and accompanying text. If the evidence DNA and suspect DNA are matched at different loci than those that led to the data bank match, the probative value of the DNA evidence can be substantial.

40 A special problem is posed by the fact that most DNA data registry proposals seek to include only those with criminal records, often for types of crimes, like rape, in which DNA evidence is likely to be used. The criminal record is itself likely to figure in some way as evidence against the accused (even if it leads only to spoliation inferences from the defendant's failure to take the stand for fear that the past conviction would be revealed), and it too is a reason why the defendant was eligible for suspicion. Putting aside civil liberties and cost concerns, a good case can be made that if DNA data bases are to be maintained, they should not
A. Implications of the Likelihood Ratio

Turning now to the likelihood ratio, we see that the probative value of DNA evidence depends on the ratio of the probability that we would find the DNA evidence if the defendant were the source of the DNA, to the probability that we would find that same evidence if the defendant were not the source. If, for example, we had examined a monomorphic allele, the ratio would be one, and the evidence, as our intuition tells us, would have no probative value and so would be irrelevant in deciding whether the defendant were the source of the evidence DNA. When a polymorphic allele is examined and there is a match, the likelihood ratio will exceed one, thus increasing the odds we would place on the defendant’s guilt. The degree to which the odds are increased depends on the frequency with which the specific allele is found in the relevant suspect population. It is this case, that of a matching polymorphic allele, which I shall focus on here.

First, notice that while it is convenient to think that if the defendant is the source of the evidence sample the probability of a match is one, this is not precisely the case because the probability in question is not the probability that the defendant’s DNA is the same as the DNA in the evidence sample but is instead the probability that a match will be reported. However, given that a match has been reported, we shall for expository purposes treat this probability for the moment as one and focus our attention on the denominator, the probability that a match will be reported if the defendant is not the source of the evidence DNA.

Our first estimate of this latter probability is the population base rate frequency for the combination of alleles in question. The base rate tells us the probability that a person randomly selected from the population will have the same DNA profile as that of the evidence sample and the defendant. One must, however, define the population from which a hypothetical person should be randomly chosen. DNA laboratories currently use convenience samples which they regard as representative, with respect to the alleles probed, of Caucasians, Blacks, or Hispanics generally. But the appropriate population is, as we have seen, not the general population but the population of possi-

---

41 A monomorphic allele is an allele that does not vary from person to person. Thus the fact that an evidence sample and a suspect sample share such an allele does not serve to distinguish the suspect from anyone else who might have committed the crime.

42 The FBI laboratory’s Caucasian data base was, for example, originally based on the analysis of blood samples taken from FBI agents, and other laboratories have typed blood samples from blood banks to establish their reference data bases.
ble suspects. This will often mean a population composed of people who are of the same race as the defendant, who are of the same gender as the defendant, and who live in the same area as the defendant. The use of generalized population data bases are, strictly speaking, proper only if it is reasonable to suppose that, with respect to the alleles probed, both the suspect population and the sample that makes up the data base are essentially random subsets of the same larger population. If this is not the case, the expert will miscalculate the probative value of a match. I have already pointed to the problems posed by both population substructure and the likely presence of relatives in the suspect population. These problems mean that when current data bases are used to determine the denominator of the likelihood ratio, the danger of misleading the jury cannot be dismissed.

The denominator of the likelihood ratio also depends on the standard that is used to declare a match. The more generous the standard, the more likely it is that a match will be declared even when the defendant is not the source of the DNA. If the standard used to declare a match (e.g., within $N$ standard deviations) is not the same as the standard used to categorize allele frequency in the population data base, use of the data base figures will be inappropriate and may distort the likelihood ratio by an order of magnitude.

The Bayesian model suggests that as a general matter the population base rate, even if accurate, is likely to be a somewhat low estimate of the probability that evidence of a match would exist if the defendant were not the source of the evidence DNA. This is because in certain circumstances a match would be reported where a match was not apparent; namely where nonmatching bands were attributed to band shifting or other analytic problems. Given that a match

43 See supra notes 9-21 and accompanying text.
44 But see studies cited infra note 76. It may well be that a match is so improbable that an exaggerated picture of its improbability caused by an inappropriate data base does not prejudice a defendant because the true probability would have been more than enough to persuade a jury that the defendant was the source of the evidence sample. Moreover, even if the defendant belongs to an ethnic group with alleles distributed quite differently from those in the laboratory's reference data base, this is unlikely to lead to a serious underestimate of the probability that a random person left the evidence DNA unless a number of individuals belonging to the same ethnic group are also members of the suspect population.
45 One might think that no forensic laboratory would make such a mistake where a person's life or liberty was at stake, but it has happened. See Lander, DNA Fingerprinting on Trial, 339 Nature 501 (1989).
46 Remember it is the reporting of a match, not the fact of a match, that is the evidence in the case. The reporting is only fallible evidence of the fact of a match.
47 Band shifting occurs when DNA fragments in one lane of a gel migrate at a different rate than identical fragments in other lanes of the same gel. Thus the DNA from the evidence sample may have moved a greater distance than the DNA in the suspect sample even though
has been reported, unless it is absolutely certain that a particular distortion of a true match has occurred, the probability of the reported match if the defendant is not the source of the evidence sample is no longer the frequency of the configuration of matched alleles in the population. It is that frequency plus the frequency of the configuration of alleles apparently found and the frequency of any other alleles which, had they been found, would have been interpreted as a match.\(^49\) This number may be difficult or impossible to estimate and is perhaps best treated as a further reason for conservatism in estimat-

they have the same source. Where bands in the evidence and suspect lanes would match but for a degree of displacement common to the alleles probed, forensic experts have sometimes called a match or declared a marked difference in migration inconclusive despite the apparent difference in the size of the evidence and suspect alleles. There are ways to check for band shifting by including monomorphic alleles, which are known to be the same size across persons, in the analysis.\(^48\)

For an excellent discussion of those aspects of DNA analysis that may make the existence of a match ambiguous see Thompson & Ford, *The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints*, in *FORENSIC DNA TECH.* ch. 7 (M. Farley & J. Harrington eds. 1991).

In other words if a match would be reported if two alleles matched precisely or if each allele in the evidence lane of the gel were displaced precisely the same distance from the alleles in the suspect lane, the probability of a match being reported if the defendant were not the source of the suspect DNA would be the probability that a random person would have alleles precisely matching the evidence DNA plus the probability that he would have any two alleles measurable by the test, and displaced the same amount give or take measurement error, from the alleles in the evidence DNA. Moreover, to pursue still further the implications of allowing claims of matches in situations like band shifting where there is displacement of the DNA across lanes, an implication of the Bayesian analysis is that even when there is a precise match the frequency associated with the match should be the frequency of the allele in question in an appropriate population plus the frequency of all other alleles that would have led the testifying expert to call a match despite disparities. This is the case because before the test is conducted this is the probability that the expert will report matching alleles if the suspect did not leave the evidence sample. For this reason discrepancies attributed to band shifting or contamination should never be regarded as more than inconclusive unless scientifically valid procedures indicate conclusively that displacement in precisely the degree observed has occurred. Indeed, except to counter a spoliation inference (e.g., a jury belief that if DNA evidence was not presented by the state there must have been an exclusion), the Bayesian model suggests that DNA testimony should not be admitted where results are inconclusive unless the expert can give the jury a reliable estimate of the relative probabilities of exclusion and inclusion. If such probabilities cannot be given, and at the moment they cannot, a jury might well believe that inconclusive means that there is a fifty-fifty chance that the defendant left the evidence DNA. Such a belief can have substantial incriminatory value.

Note finally that the numerator of the likelihood ratio is also affected when there is no actual match between the evidence and the suspect samples and the deviations from a match are attributed to analytic or test problems. In these circumstances unless absolute confidence in the explanation for the deviation exists (i.e., the conditions of the analysis are such that one knows for certain that if the suspect and evidence DNA are the same there would be the deviation between the two samples that was found), the probability that there would be a "match" with the deviations found if the defendant were the source of the evidence DNA will be less than one, thus lowering the likelihood ratio even without taking into account denominator effects.
ing allele frequencies in the first instance and as a reason not to call a match unless a match is apparent.

B. The Implications of Error

It is obvious that both the numerator and denominator of the likelihood ratio may be affected by any laboratory procedures that allow for error, with the result that the likelihood ratio, which is to say the probative value of the evidence, will diminish. Possible sources of error include sloppy laboratory procedures, the materials used, the quality of the evidence DNA, and the protocols used for calling a match. Moreover, any information that influences the calling of a match apart from the physical evidence of the autorads can have a similar effect. Thus laboratory failures may lead to both mistaken exclusions and false positives. Other kinds of failures may pose a threat of only one kind of error. Thus if an analyst is influenced in calling a match by knowing whether it is a prosecutor or defense counsel who has submitted the sample, false positives or mistaken exclusions, but not both, would become more likely.

What may be less obvious is how serious the implications of error are for the kinds of probability statements that are often given by prosecution experts in DNA cases. Recall that in the Bayesian model as presented the numerator of the likelihood ratio was \( P(E|S) \) and the denominator was \( P(E|not-S) \), in which \( E \) stood for evidence of a match between the evidence and suspect samples and \( S \) was the fact that the defendant was the source of the DNA evidence sample. But it is not the fact of a DNA match that is presented at trial; rather it is testimony that a DNA match in fact exists. Yet witnesses testifying about the probability of \( (E|not-S) \), which is to say the probability that the DNA came from someone other than the defendant, present figures that are based not on the likelihood that a match would be reported if the defendant were not the source of the evidence sample but rather on the likelihood that the DNA would in fact match if the defendant were not the source of the evidence sample. The latter figure can be tiny indeed; figures smaller than one in fifty billion have been presented at trials. The former figure, that is the likelihood

---

50 For a review of these sources of error, see Thompson & Ford, supra note 3; Thompson & Ford, supra note 48.

51 The possibility of such influences is an important reason why visual matching should not by itself be sufficient to call a match. There should be some objective check on visual matches or, indeed, objective matching in the first instance. Current "objective" systems allow the analyst to override the machine settings.

52 Cf. Lempert, supra note 33, at 1052-55.

53 Note that critiques like those made above, see supra text accompanying notes 11-25, and
that a match would be reported if the defendant were not the source of the evidence sample, can never be less than the false positive rate of the laboratory examining the DNA in question.

Forensic experts often present their findings with great confidence, but infallibility is unfortunately not a characteristic of forensic laboratories. In one study in which samples of known materials were sent to a number of forensic laboratories, error rates were as high as seventy percent! Another study reports crime laboratory error rates as high as seventy-seven percent. While labs doing DNA testing have done better than this in the limited blind proficiency testing that has occurred, they are still far from perfect. In the only blind DNA tests reported in the literature, the three major commercial laboratories then doing forensic DNA testing were sent fifty samples each. Two of the three declared one false match. In a second set of tests the year following, one company declared a false match. Nor have false positive errors been confined to tests in which no one's fate was at stake. Eric Lander, for example, describes a case in which one of the major commercial laboratories reported that a woman was the mother of an apparently abandoned newborn; later it turned out that the woman was pregnant with another child at the time the abandoned newborn was discovered. In another case a second major commercial laboratory compared two blood samples from the same alleged rapist and declared they did not match, thus supporting the prosecutor's contention that in providing the first blood sample,

by others who have examined the forensic use of DNA suggest that such figures are far smaller than can be justified (see, e.g., Thompson & Ford, supra note 3; Lander, supra note 9), but the argument here proceeds on the assumption that such figures are accurate estimates of the likelihood of a purely coincidental match.

56 OFFICE OF TECHNOLOGY ASSESSMENT, U.S. CONGRESS, GENETIC WITNESS FORENSIC USES OF DNA TESTS 79 (July 1990) [hereinafter GENETIC WITNESS] (interview with B. Grunbaum and C. Moraga). Other studies in this report quote more modest error rates, in one case as low as 2.4%, or 1.3% if the laboratories making more than three errors are eliminated. Some of the discrepancy in error rates between studies is apparently attributable to how error is defined, for example, whether inconclusive results are defined as error.
57 The tests were not fully blind. While the laboratories did not know whether samples sent matched, they did know these were test samples. The one lab that declared no false matches on the test, apparently had its researchers run the DNA tests rather than the technicians that usually performed them. Note, DNA Profiling, supra note 3, at 493. Another lab was allowed to withdraw a report with several errors and submit one with only one error after its representative met with the testing committee. Thompson & Ford, supra note 48, at 144 n.122.
58 GENETIC WITNESS, supra note 56, at 79-80.
59 Lander, supra note 45, at 505.
which resulted in an exclusion, the suspect had substituted a third party's blood for his own. It took a third test, insisted on by the suspect, to show that the second laboratory had made a mistake and that he could not have been the rapist. 60

Yet analysts from the same labs that made these errors (perhaps the same analysts) will report matches and confidently tell the jury that the probability of such a match if the defendant was not the source of the evidence sample is less than one in many billion. In fact, to judge by the proficiency test results, the probability that the jury would be told of such a match if the defendant were not the source of the sample is closer to one in fifty or one in a hundred.

The probability of a reported match if the suspect did not leave the evidence DNA is greater than a laboratory's false positive rate because in addition to the possibility of error there is also the possibility that the DNA was left by someone other than the defendant. However, the latter probability is usually dwarfed by the probability of a false positive error; so this apparently crucial probability, which is at the center of most disputes about the admissibility of DNA testimony, may ordinarily be safely ignored! Indeed, jurors provided with a laboratory's false positive rate and with information about the likelihood, assuming no testing error, of a match if the evidence DNA was not the defendant's, are likely to be hopelessly confused about the weight to accord the testimony because ordinary people are not very good at working with conditional probabilities. 61 Thus jurors ordinarily should receive only the laboratory's false positive rate as an estimate of the likelihood that the evidence DNA did not come from the defendant. 62

---

60 Id.
61 See, e.g., Schum & Martin, supra note 32.
62 Only when the characteristics of the evidence DNA are so common that the likelihood it could have been left by a random individual is close to or exceeds the false positive rate should jurors be informed of this probability because only then will the probability of matching a random member of the suspect population be sufficiently close to the error rate that knowing this rate will appreciably increase the likelihood of a false match beyond what it would appear to be from knowledge of the false positive rate alone.

Where the random match probability is of this magnitude, the statistic should not be presented to the jury alone. Rather it should be added to the false positive rate. The resulting statistic will inform the jurors of the likelihood that a DNA match would be reported when the defendant was not the source of the evidence DNA. It is this statistic that specifies the weight that should be given the DNA evidence.

After a draft of this section of the article was completed, I learned that my basic point about the implications of laboratory error rates had been made by Paul Hagerman in a letter to the American Journal of Human Genetics, Hagerman, DNA Typing in the Forensic Arena, 47 AM. J. HUM. GENETICS 876 (1990). Hagerman's argument was implicitly accepted in another letter to that journal by Russel Higuchi, a scientist who works for one of the labs that has been a pioneer in forensic DNA testing. R. Higuchi, Human Error in Forensic DNA Typing, 48
Estimating the appropriate false positive rate for a DNA test is no simple matter.\textsuperscript{63} Ideally, false positive rates should be specific to the laboratory, the analyst, and the quality of the DNA examined.\textsuperscript{64} We cannot get such data. Typically the best evidence of the likely laboratory error rate is the error rate revealed by blind proficiency testing, which might be thought of as an average error rate over a nonrandom subsample of the kinds of cases the laboratory is likely to encounter. Moreover, even at the laboratory level, blind proficiency testing might be inadequate to estimate risks of error.\textsuperscript{65} DNA testing is expensive and time consuming; thus each lab is likely to be subject to a relatively small number of blind tests in a given year. If a laboratory error rate is, for example, 1 in 500, which would reflect extraordinary accuracy in most kinds of forensic analysis, the error rate is still likely to be orders of magnitude larger than the frequency estimates that the laboratory would put on the possibility that a match would have been reported had the DNA come from someone other than the defendant. Yet with an error rate as low as 1 in 500, many

\textit{AM. J. HUM. GENETICS} 1215 (1991). Hagerman demonstrates that the probability of a false match is essentially independent of the population frequency when the laboratory error rate is much larger than the frequency of a given band pattern in the general population. Contrary to my argument in the text, Hagerman suggests that the laboratory false positive rate be added to the estimated frequency rate as a matter of routine. Hagerman concurs with my suggestion in the text at note 64 \textit{infra} that DNA samples should be split and independently analyzed. He also cautions, as I do in note 63 \textit{infra}, against the danger of mistaken exclusions.

\textsuperscript{63} I am focusing on false positive rates because these rates are at issue in determining the probative value of the DNA evidence that prosecutors seek to introduce. However, prosecutors deciding on whether to drop cases or judges and jurors evaluating evidence of a DNA exclusion must concern themselves with false negative rates, and these are probably higher than the incidence of false positives. False negative rates are a special concern because if the DNA from two sources differs, random analytical errors are unlikely to result in the appearance of a match, but if the DNA from two sources is the same, random errors caused by degradation or analytical mistakes are likely to result in the appearance of a nonmatch.

\textsuperscript{64} Laboratories vary with respect to such matters as the loci they probe, the quality of their statistical data bases, the procedures and standards they use to declare a match, and the care that is taken to prevent the contamination of one sample of DNA with another. Analysts vary in their skill and in the way they resolve ambiguous cases. Some evidence samples contain DNA in large quantities in good condition while other samples are smaller or degraded and contaminated in various ways. And some alleles that do not match are close in size while others are quite different. The interaction of these factors means that the likelihood of false positives cannot be expected to be constant across laboratories or within laboratories across analysts and cases analyzed.

\textsuperscript{65} Even conducting a truly blind proficiency test may pose difficulties. If, for example, DNA samples are typically sent to laboratories by the police along with summaries of other evidence in the case, test samples should include summaries which report incriminating information with about the same detail that police summaries usually do. If DNA samples contain the number of a contact person who can be called for further information about the case, blind samples should also give the number of a contact person who must be available to discuss the "case" if called. Test samples should also be degraded in varying degrees to match the degrees of degradation that are found in actual casework.
years might pass before an error occurred in a blind test. Thus unless forensic laboratories have uncomfortably high error rates, the primary virtue of blind proficiency testing is not that it allows error rates to be estimated. Rather, this virtue is that knowing that any sample analyzed may be a test is an incentive for quality control in all tests the laboratory runs.

One response to the possibility of laboratory error and the limits it should place on the weight accorded DNA matches is to standardize and improve technology and training to lower error rates substantially. This is of course desirable, but so long as humans are involved in the identification process, it is utopian to expect error rates that will not dwarf the probabilities that could be associated with reported DNA matches in an error-free world. Thus in this world, proficiency testing, licensing and other such palliatives are not enough. Instead, whenever possible DNA should be sent to two or, better yet, three laboratories for independent analysis. Since most of the sources of false positives are independent across laboratories, laboratory error rates could be multiplied to give the possibility that all laboratories reporting a match are mistaken, and either this figure or the probability of a random match, whichever is greater, could be given to the jury as the probability that a match would be reported when someone other than the defendant left the evidence sample. If, for example, each of three testing laboratories had a false positive rate of 1 in 100 the probability that three laboratories would all mistakenly call a match is about one in one million. It is expensive to send DNA to two or three labs for separate analysis, but if prosecutors wish to present jurors with the overwhelming numbers that make

---

66 An error rate of 1 in 500 is not small when the failure to recognize this will lead a jury to conclude that the probability of a mistaken identification is less than one in a billion. Yet it is small if the issue is the likelihood that an error will occur in a blind proficiency test when a lab is tested on, for example, only 10 or 20 samples a year.

67 From this perspective it is a good rather than a bad thing that the leading laboratories use different probes and have different data bases.

68 If the figures were similar in magnitude, error from both possibilities should be considered in setting a probability for the jury. This can be done, as Hagerman, supra note 62, points out by summing the two probabilities and reporting the sum to the jury.

Note that the probability of a false positive and the probability that a random person left the evidence DNA are not the same in their implications. The latter probability but not the former has implications for estimating the number of members of the suspect population who might have left the evidence DNA. If the false positive rate is thought to be the same as the "random match" probability there may be a particular danger of making the "defense attorney's fallacy." Thompson & Schumann, supra note 7, at 171.

69 Actually, the combined false positive rate is probably somewhat greater because certain sources of error might inhere in the sample or be common to all laboratories.
prosecutors so enamored of DNA evidence, either laboratories must demonstrate they perform virtually without error or multiple testing must occur.

C. The Value of A One Allele Match

One final implication of the Bayesian model is that the probative value of even limited evidence may be surprisingly large. Thus in a situation where only one allele is tested, if there is a match between the evidence DNA and the defendant's DNA and if the population base rate for the matched allele is accurately assessed at 1 in 100, the evidence would mean that a juror's prior odds on guilt should be multiplied 100 fold after receipt of the DNA evidence. This could easily change a situation in which the apparent balance of the probabilities went from one which suggested that the defendant was not guilty, to odds that justified the conclusion that the defendant was guilty beyond a reasonable doubt. Yet our intuition suggests that on the basis of one match this would be an unreasonable leap.

There are two reasons why our intuition may be correct even though the Bayesian implications are not wrong. The first is that as a psychological matter jurors are likely to have unreasonably high prior odds. Thus people who do not know whether or not a defendant is guilty often give initial odds on guilt of 1:1 or fifty-fifty. In fact the odds should be one to whatever number of people make up the suspect population. Thus, multiplying prior odds, properly understood, by a number as small as one hundred is still likely to leave it far more likely than not that reasonable doubt of the defendant's guilt remains.

Second, there is the so-called spoliation issue; that is the implications of what is absent from evidence. If a person were only presented with the information that one of the defendant's alleles matched one allele in the evidence DNA, she would wonder about the other alleles

---

70 Subject to the caveats about micro populations, see supra notes 11-27 and accompanying text.
71 In the context of DNA evidence, virtually without error means that error rates must be the same low order of magnitude as the frequencies associated with the probability of a random match.
72 To some extent the more alleles tested by a single laboratory, the more confident we can be of the accuracy of a reported match across all alleles. Certain errors might affect only the measurement of certain alleles, and the fact that a number of other alleles match may be such good evidence that the suspect and evidence samples came from the same source that we are justified in disregarding the possibility that the claim that a particular pair of alleles match is in error. Other mistakes, however, such as contamination of the evidence lane with suspect DNA, will yield false matches no matter how many alleles are tested. Until large numbers of blind proficiency tests occur without error and proven errors cease to be found in tried cases, we must assume error rates many orders of magnitude larger than the random match probabilities often reported by DNA analysts.
tested and why evidence about them had not been presented. The suspicion would be that evidence of the other alleles was exonerative or that there was a fear that if further testing was conducted the results would be exonerative. If this concern is reasonable, the conclusion that one matching allele does not justify increasing the estimated prior odds on guilt one hundredfold would be justified unless some good reason could be given why it was impossible to test for more than one allele. If it were impossible to test for more than one allele and the one allele that could be tested for did match, this would be relevant evidence, and so long as it was unlikely to confuse a jury, it should be admitted.

D. The Number Of Alleles To Test

Closely related to the issue of the evidential implications of one-allele matches is the issue of how many alleles should be tested to determine whether evidence and suspect DNA match.\(^73\) Currently it appears that three or four loci which may yield as many as six or eight different alleles are ordinarily probed. The adequacy of this number depends not only on the population frequencies of the alleles identified but also on the validity of the assumptions of Hardy-Weinberg equilibrium\(^74\) and linkage equilibrium\(^75\) and on the costs of erroneous judgments. Not only are the equilibrium assumptions questionable,\(^76\)

\(^73\) The discussion that follows assumes that my suggestion that the DNA probability evidence be limited by false positive rates will not be followed. If it is, the error rate will impose a limit on the probability associated with a reported match that cannot be substantially lowered by testing more alleles. (But see supra note 72.) However, testing more alleles may still be justified in that it gives a higher probability of finding true exclusions. Of course the probability of reporting a false exclusion increases as well.

\(^74\) The Hardy-Weinberg equilibrium assumption justifies treating the two alleles a person possesses at each locus as independently distributed for the purpose of determining their joint probability.

\(^75\) The linkage-equilibrium assumption justifies treating alleles at different loci on different chromosomes as independently distributed for purposes of determining their joint probability, or so the term is used in much of the writing on the forensic use of DNA. However, in the field of genetics, linkage, strictly speaking, refers to the association of genes on a single chromosome. See definition of linkage disequilibrium in R. King & W. Stansfield, A Dictionary of Genetics 181 (4th ed. 1990). Population substructure, for reasons apart from linkage, may mean that within a larger population the assumption that alleles on different chromosomes are independently distributed will not hold. See, e.g., Walsh & Buckelton, supra, note 21.

\(^76\) See the sources cited supra at notes 3, 9. There is considerable dispute over how robust estimates of DNA band pattern frequencies are to the empirical violation of these equilibrium assumptions. Contrast the articles noted above with Budowle, Giusti, Waye, Baechtel, Fourney, Adams, Presley, Deadman & Monson, Fixed-Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci, for Use in Forensic Comparisons, 48 Am. J. Hum. Genetics 841 (1991) [hereinafter Budowle]; Devlin, Risch & Roeder, No Excess of Homozygosity at Loci Used for DNA Fingerprint, 249 Sci. 1416 (1990); Evett & Gill,
but testing laboratories do not ordinarily adjust the number of matches they test for on the basis of the costs of false positives. A good case can be made that such adjustment should occur, and that at least where an identification that might lead to the death penalty is involved an additional locus or two should be tested because of the potential costs of error. Indeed, it might make sense to establish a standard that provides that for DNA evidence to be admissible, if the amount of available DNA permits, a minimum number of probes must be used and a certain low probability of random inclusion must be reached. The lower the probability of random inclusion, the less difficulty a jury should have in weighing the information. This too might argue for the testing of more alleles than is conventionally done.

E. Presenting DNA Evidence

Finally, we come to the question of whether the statistics of DNA evidence should be presented in Bayesian rather than frequentist terms. Finkelstein and Fairley have argued that statistically-based identification evidence should as a general matter be presented in a Bayesian format, and Evett and Werrett have extended this argument to DNA identification evidence in particular. Many conventionally trained forensic scientists would, no doubt, argue against this proposition.

One common argument against—that juries would have difficulties in understanding Bayesian presentations—may be less substantial than it appears. Jurors have difficulties understanding any statistical
presentation of evidence, and to date there is no good evidence which suggests they have more trouble understanding the implications of Bayesian than of frequentist perspectives. Indeed, since the Bayesian presentation is directed to the question the jury must resolve, jurors might find a Bayesian presentation of DNA statistics more intuitively comprehensible than a frequentist statement.

The need to estimate prior odds to evaluate the probative value of the evidence presents, however, a more formidable problem. Evett and Werrett seek to avoid the problem by arguing for a convention whereby a jury would be told that the evidence weakly supports a finding of identity when the likelihood ratio is no greater than ten, that it supports the finding when the ratio is between 10 and 100, that it strongly supports the conclusion when the ratio is between 100 and 1,000 and that it very strongly supports it when the ratio is greater than 1,000. However, this and similar conventions can be misleading because the probative implications of the likelihood ratio is inextricably linked, as we have seen, with the estimated prior odds. If the prior odds are very small, even likelihood ratios of 1,000 or more may yield posterior odds on identity that are substantially below fifty-fifty. Thus if jurors are to be given Bayesian likelihood ratios, they should also be instructed in how they might estimate prior odds and integrate their judgments of prior odds with the probative force of the DNA evidence as represented by its likelihood ratio. Whether this task could be easily accomplished is a question that I cannot answer, but the possibility that DNA identification evidence could be presented more effectively in a Bayesian than a frequentist fashion deserves further exploration. Alternatively, if other evidence suggests that reasonable jurors should perceive at least a moderate prior probability of guilt, a convention like Evett and Werret’s might well be the best way to proceed.

III. HOW DOES DNA EVIDENCE MATTER?

Despite all the fuss that has been made about DNA evidence, no

---

81 For a review of studies dealing with the competence of jurors to evaluate statistical evidence, see Thompson, Are Juries Competent to Evaluate Statistical Evidence? 52 LAW & CONTEMP. PROBS. 9 (Autumn 1989).

82 Evett’s thinking about the appropriate verbal labels for different likelihood ratios has changed over time. For a more recent view see I. Evett, supra note 30. The precise numbers and labels do not matter here, for the scheme is only offered by way of example.

83 So long as the likelihood ratios for each allele are conditionally independent of the ratios for each other allele, they may serve in turn as multipliers of the prior odds. Thus the likelihood ratios may in theory be presented sequentially or as a composite for N matches. However, it is better practice to present the composite figure, for jurors are not experts at combining probabilities. See Schum & Martin, supra note 32.
one knows how much it matters. While there are now numerous cases in which DNA evidence has been introduced, this does not mean DNA evidence matters greatly. It is important to recognize that there are two uses to which forensic identity evidence may be put. First, the evidence may be used to identify an otherwise unknown criminal. Second, the evidence may be used to prove that a defendant is in fact guilty. It appears that evidence which is presented to the jury as if it were crucial identification evidence is often important primarily as a way of proving a defendant guilty.\footnote{For a particularly perceptive article, that states and builds on this insight, see Gross, \textit{Loss of Innocence: Eyewitness Identification and Proof of Guilt}, 16 J. of Legal Stud. 395 (1987).} Police line-ups, for example, are often conducted after the police are confident they have collared the suspect.\footnote{Id.} The lineup, although it seems to be a crucial stage in the identification process, may in fact be staged largely for its dramaturgical value.\footnote{This is not to say that a victim's failure to identify a criminal at a lineup will necessarily have no effect. At a minimum it is likely to lead the police and prosecutor to reevaluate the "goods" they have on the suspect.} The prosecutor, in other words, may be able to present a more persuasive case with a lineup identification than without one.

The situation may be similar with DNA; after all one must first have a suspect before a DNA test can be ordered.\footnote{But see the English case of \textit{Regina v. Pitchfork} described supra note 18.} In how many of these cases would the evidence that led to suspicion have been sufficient to lead a jury to convict, absent the DNA evidence? We simply do not know. Yet the distinction is important in evaluating the danger that DNA evidence will lead to the conviction of the innocent. If DNA evidence were used largely as a dramaturgical device in cases that were otherwise so strong that a conviction should occur without the DNA identification, there would be little danger to the innocent from this new technology. In terms of the Bayesian model, the prior probability of guilt (i.e., the probability based on other evidence in the case) would be so high that despite the dangers of overweighing inculpatory DNA evidence there would be little prospect that overweighing would lead to injustice. If, for example, the odds on defendant's guilt before the introduction of DNA evidence were one hundred to one in favor, it would hardly matter whether the odds after the introduction of the DNA evidence were a reasonable fifty thousand to one or an absurd fifty trillion to one. Either way, the defendant will be convicted, as he would have been had the DNA evidence never been offered. Indeed, one might see the DNA evidence in such a case as...
providing insurance that the jury would not misperceive the weight of the non-DNA evidence and mistakenly acquit. The situation is different, however, if the DNA evidence is essential to identify the defendant as the criminal. In this situation the DNA evidence is of real as well as dramaturgical importance, for as evidence of identity it is not redundant with other admissible evidence on this issue. Thus if a match has been mistakenly declared, a mistaken conviction is a possible and perhaps likely result. Where DNA evidence is used this way, attention to the caveats that I and others have raised is essential to avoid instances of injustice.

The problem will become particularly acute as planned offender data bases come on line, and individuals are identified as criminals on the basis of data base matches. In these circumstances there is a substantial likelihood of misidentification. First, it is easy to forget that the prior probability that the person identified is in fact the criminal is quite low. Thus, even after a DNA match has been found, it may be more likely than not that the individual identified did not commit the crime. Second, the probability that the defendant's DNA would match the evidence DNA is not the probability that

88 Mistakes might still matter in such a case, for it is conceivable that but for a mistake a DNA exclusion would have been observed, one that would justify an acquittal despite the apparently high prior odds in favor of guilt. However, the higher the prior odds in favor of guilt, the more likely it would be that a DNA exclusion was a false negative.

89 As of January 1990, at least 11 states had enacted laws to require some level of DNA typing of convicted offenders and in other states databanking legislation had been proposed but was not yet enacted. GENETIC WITNESS, supra note 56, at 122-24. A LEXIS search conducted on August 23, 1991, revealed an additional eight states. (library: CODES; files: ALLCDE & ALLALS; search term: "DNA").

90 The prior probability would be one over whatever number constitutes the suspect population. The exact figure would reflect the increased likelihood that a person convicted of the same crime(s) as the suspect would, as compared to a person who has not been convicted of these crimes, commit the crime in question, and, if the analysis is to be conducted at this level of precision, an adjustment that takes account of the number of other people in the suspect population who have been convicted of crimes and of their increased likelihood of committing the crime in question. This figure cannot be calculated—we lack the needed information—but one may contemplate its likely magnitude for heuristic purposes.

Note also that when an identification occurs from a databank match, the suspect population may be quite a bit larger than it ordinarily is. This is because a person whose DNA matches evidence DNA will be treated as a suspect even if he lives so far from the crime that he would not ordinarily have been considered a member of the suspect population. If it is credible to believe that a suspect living at such a distance committed the crime, then all males of an appropriate age (assuming a rape) who live within that distance should be considered as potential members of the suspect population.

91 It may, however, be difficult for the suspect to convince a jury that the identification is wrong, since if he testifies, and perhaps even if he does not, the jurors will learn of the crimes that caused his DNA to be placed in the databank. The jurors are likely to be unaware of the fact that given the way he was identified the defendant had to have had a criminal record, and they might find this record unduly probative of guilt.
would ordinarily be associated with the DNA match. Suppose, for example, that the probability that a random member of the suspect population would possess DNA matching the evidence sample is one in a hundred thousand. If a suspect's DNA were to match the evidence DNA, the reported match would seem to be highly incriminating. But if the suspect was identified by scanning a data base that contained information on twenty-thousand people, finding a match in this data base, regardless of guilt, would not be surprising since there are twenty-thousand chances that a match might be found.

Thus when a suspect is identified by matching evidence DNA to a data base, if DNA identification evidence is to be presented at trial, the alleles reported on should not be those originally matched but should instead be a set of alleles located at different loci. If these too match, the evidence is likely to be highly and properly incriminating. But even in these circumstances one can have more confidence in the suspect's guilt the greater the state's ability to provide other evidence that suggests guilt. In particular, discrepant evidence, such as the fact that the suspect does not resemble the victim's original description, should not be lightly dismissed because it seems insubstantial when contrasted with apparently powerful DNA evidence.

**CONCLUSION**

I conclude where I began. DNA evidence is a valuable tool for solving crimes, especially rape, and for avoiding erroneous convictions. At the same time the technology for DNA identification is in its infancy, and even as this technology matures, laboratories will never be perfect. Thus there are caveats which counsel against overly heavy reliance on the results of DNA testing. In particular, in cases where the suspect and evidence samples match, the cautions that I and others have raised counsel against presenting extremely low estimates of the probability that the evidence DNA might have come from someone other than the defendant. Even if forensic scientists in this country do not follow Evett's lead in analyzing the probative value of DNA matches in a Bayesian framework, they might be wise to follow his lead and translate their probabilities into the simple English terms weak, fair, good, strong, and very strong, and leave it at that. Given laboratory error rates, micro population considerations, and the like, a good case can be made that once the jury has been told that

92 See *supra* notes 30, 80. It should be pointed out, however, that Evett's Bayesian approach and his verbal conventions have to date been used only for research and not for casework purposes. The Home Office's routine casework procedure is still match/binning with the frequentist probabilities that go with it.
a DNA match is very strong evidence of guilt, there is no more that can be responsibly said. This limitation should not be viewed with despair by prosecutors or by citizens who wish to see criminals punished. Currently we convict most of those charged with serious crimes. With very strong DNA evidence, conviction rates can only increase.

A POSTSCRIPT ON LEGAL ISSUES

A. Frye or "Helpfulness"

I have talked in this paper only about the proper evaluation of DNA evidence. By way of a postscript I shall discuss briefly two legal issues that have confronted courts deciding whether to admit DNA evidence. The first concerns the question of the standard the court should apply. The two contenders are the Frye test and the so-called "helpfulness" test of FRE 702 and its state counterparts. Under Frye, to be admissible, a new technique "must be sufficiently established to have gained general acceptance in the particular field in which it belongs."94 Under FRE 702 an expert opinion should be admitted when "scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue . . . ."95

Courts and commentators often speak of these tests as if they are different, with the Rule 702 test being more favorable to admissibility. My view is that when an issue like the admissibility of DNA evidence is first raised the two tests have, or at least should be interpreted to have, the same meaning, a meaning close to what the Frye court expressed. When the probative value of evidence turns on questions concerning the validity of theories, the adequacy of technologies and the interpretation of data that are on the cutting edge of several modern sciences, scientifically naive judges are not well-equipped to determine by themselves what is from a scientific standpoint sufficiently sound to be relied upon. Nor are judges likely to be able to decide these matters correctly on the basis of testimony from competing expert witnesses. A solution judges sometimes adopt in this situation, which is to treat the disagreement between experts as going to weight rather than admissibility, has little to commend it. Jurors are likely to

93 FED. R. EVID. 702.
94 Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).
95 FED. R. EVID. 702 (emphasis added). Note that a revision of this rule has been proposed by the Advisory Committee on Civil Rules to the Committee on Rules of Practice and Procedure of the Judicial Conference of the United States. It is currently at the "notice and comment" stage.
be in no better position than a judge to decide rationally between competing scientific views. Rather they are likely to be confused, which is a reason to exclude evidence under FRE 403.96

What courts should do, even under FRE 702, is what Frye seems to command.97 They should seek to determine the general consensus in the relevant scientific community, for it is only this group of scientists that is well-equipped to assess the validity of the theories and procedures on which the probative implications of the evidence depends. The consensus required to justify novel scientific evidence need not, however, be universal, for there are often dissenters, even reputable dissenters, to what is scientifically valid.98

While the law must admit the possibility that what most scientists "know" is wrong, in deciding what is scientifically sound a court usually can do no better than proceed on the assumption that what scientists generally know is correct. But if there is substantial disensus within the relevant scientific community, or if in a certain sphere the scientists who are most expert in a matter disagree with the conclusions of other scientists, the evidence should be excluded for the time being. Given the general adequacy of trial procedures at any particular point in time, we should not risk overwhelming the jury with information whose validity rests on propositions that are questioned by many knowledgeable scientists. If the skeptical scientists are wrong, a consensus regarding where the truth lies should emerge soon enough, and little harm is likely to occur in the meantime. Certainly the situation is not likely to be worse than the status quo ante, a status quo that probably seemed satisfactory until the prospects of a new scientific technique emerged. To hear some proponents sing the praises of DNA, one would think that no rapists (except perhaps innocent ones) had been convicted before 1987. It ain't so.

Specifically with regard to DNA, it appears that the most important scientific underpinnings of DNA identification procedures pass this proposed test. In particular there is a scientific consensus that no two individuals except identical twins have identical DNA, and there is a consensus that, within some margin of error, existing techniques,
like the RFLP process, can determine whether DNA from two sources matches.\(^9\) Where the consensus breaks down is on the adequacy of current data bases for determining allele frequencies in relevant populations and on the validity of the equilibrium assumptions that justify multiplying allele frequencies together to get a frequency for their joint distribution.\(^10\) Responding to these uncertainties, a few courts have excluded evidence of DNA matches because of problems with the population statistics or have admitted evidence of DNA matches but have prevented experts from using the product rule and equilibrium assumptions to attach extremely low probabilities to them.\(^10\)

### B. Weight or Admissibility

This brings me to my second point. Many courts, while recognizing that there is no scientific consensus about the data and assumptions that underlie the estimates of allele frequencies that forensic scientists present to juries, nevertheless conclude that the product rule

\(^9\) There can never be a consensus that in a given case, techniques were accurately applied because scientists as a community have no specific case knowledge. Thus, in cases where DNA evidence is introduced, it is always open to the party contesting the evidence to claim that the test was not properly run or that the test results were not properly interpreted. The textual discussion is not intended to treat the scientific issues that can arise out of facts specific to cases and so does not address the question of when testing deficiencies should be treated as going to admissibility rather than to weight. I would suggest, however, that when a testing deficiency is more likely than not to exist and when the scientific consensus is that such a deficiency, if it existed, would invalidate the test results, such a deficiency should be seen as going to the admissibility of the results in question. It is also possible in such circumstances that although a test flaw would invalidate the results reported, a more conservative estimate of the results of the test would not be invalidated. If so, the more conservative estimate should be allowed.

\(^10\) Multiplying allele frequencies together in this way is to apply the so-called product rule which holds that the probability of two events is equal to the product of their separate probabilities if the two events are independently distributed. In the case of DNA, the events are the possession of particular alleles. It is the independence of alleles and the tenability of equilibrium assumptions that posit independence which is disputed in the literature that deals with problems posed by population substructure. See sources cited supra notes 3, 9, 76. Despite these disputes it might well be possible to specify rules for counting and combining allele frequencies that are sufficiently conservative so as to satisfy most population biologists and geneticists that the frequencies they yield are a reasonable upper bound. Since such upper bound estimates could always be lowered by testing more alleles, they would ordinarily not prevent prosecutors from presenting impressively low estimates of the likelihood that a person other than the defendant was the source of evidence DNA.

may be applied to allele frequencies derived from small non-random samples to generate extraordinarily low estimates\textsuperscript{102} of the probability that someone other than the defendant left the evidence DNA and that these estimates may be given to the jury. Courts allowing such testimony treat the scientific disagreement about the propriety of applying the product rule as going to weight rather than to admissibility.\textsuperscript{103}

I believe that the courts which have decided this way are wrong. Although there is a distinction between issues of weight and issues of admissibility, and although many of the weaknesses that inher in evidence are for the jury to judge after evidence has been admitted, this is not the case with testimony reporting the weight that science would give certain results. If there is no adequate scientific foundation for the weight estimate, as evidenced by a substantial lack of consensus or a contrary consensus in the relevant scientific community, the jury cannot be assisted under FRE 702 and may be confused under FRE 403 and their state equivalents.

A hypothetical example should make this point clear. Assume a DNA analyst had a theory that no two people, except identical twins, had DNA that matched at the four loci his laboratory tested. As evidence he might show that in his laboratory's data base of 600 people there were no two people who matched at all four loci. Most scientists would agree, however, that for any current laboratory data base, allele frequency counts guarantee that if enough people are tested there will be people with the same alleles at four tested loci. This does not change the fact that this testifying scientist's claim goes only to the weight to be accorded the DNA evidence; that is, the question this testimony poses is: Should the evidence of a match be regarded as conclusive on the issue of identity or should it be given some lesser weight, such as the weight suggested by the frequencies of the matching bands? Nevertheless, it is, I hope, clear that even apart

\textsuperscript{102} There are a number of cases in which this probability has been estimated at less than one to the number of people living on earth, no doubt suggesting to the jury that the defendant must be the culprit. See, e.g., Pennell, 584 A.2d 513; State v. Horsley, 117 Idaho 920, 792 P.2d 945 (1990); State v. Lipscomb, 574 N.E.2d 1345 (Ill. App. 1991); Curnin, 409 Mass. 218, 565 N.E.2d 440; State v. Schwartz, 447 N.W.2d 422 (Minn. 1989); People v. Shi Fu Huang, 145 Misc. 2d 513; 546 N.Y.S.2d 920 (Sup. Ct. 1989); Mandujano v. State, 799 S.W.2d 318 (Tex. Ct. App. 1990).

from the possibility of error this scientist after reporting a four loci
match should not be allowed to testify that it is one hundred percent
certain that the defendant was the source of the evidence sample.

The reason why this scientist should not be allowed to testify that
his findings are conclusive of identity is the obverse of the reason why
he is allowed to testify in the first instance. He is allowed to testify to
the fact of a match (where a lay person would not be) because it has
been established that he is an expert in doing DNA analyses. If he
were not properly trained or if he did not use scientifically appropriate
methods in running his tests, he would not be allowed to testify to the
existence of a match.

One might reasonably argue that deficiencies in an alleged ex-
pert's abilities and procedures affect only the weight that should be
accorded his findings and that the jurors are perfectly able to discount
flawed findings, as well they might be. But this argument does not
meet the objection. The reason the untrained witness cannot testify as
an expert is that the standards of FRE 702 or its state equivalents are
not met. If the witness lacks appropriate training, he is not qualified
as an expert, and if he did not use scientifically appropriate methods,
the jury will not be assisted by his testimony because they will have
no basis for determining what weight his testimony should be given.
The situation is the same with respect to the hypothetical scientist's
testimony about the weight that should be given a match. Again he is
allowed to testify (where a lay person would not be) because he is
thought to have the scientific knowledge needed to evaluate the pro-
bative value of a match. When it becomes clear that his knowledge is
not regarded as scientifically sound, Rule 702's permission to testify in
the form of an opinion disappears.

It does not matter that the witness can honestly claim that his
testimony goes only to the weight to be accorded a match. Indeed, for
the prosecutor to seek the scientist's testimony, knowing it is scientifi-
cally unsupported, is akin to the presentation of evidence that the
prosecutor knows for a fact is erroneous. Even if the witness testifies
in good faith, the prosecutor has acted wrongly.\footnote{104}

There is little difference in principle between the hypothetical sci-
entist I have described and a witness who gives a purportedly scien-
tific frequency estimate that is based on equilibrium assumptions\footnote{105}

\footnote{104} For example, a prosecutor could not present a witness who, in good faith, would testify
to the defendant's involvement in a killing so similar to the crime charged as to be admissible
under FRE 404(b) if the prosecutor knows that at the time of the other killing the defendant
was in prison in a different state. The witness is not guilty of perjury, but from the jurors'
point of view he might just as well be, for they will be similarly misled.

\footnote{105} See supra notes 74-75 and accompanying text.
and assumptions about the absence of population substructure that are not generally accepted by those students of population structure and genetics who make up the relevant scientific community. If an expert's techniques for evaluating the existence or improbability of a match are not regarded as sound by most scientists in the relevant community there is no permissible basis under FRE 702 or its state counterparts for the expert's opinion. It does not matter that the expert's testimony is only directed at the weight to be given a match. In order to give an opinion at all, the expert must be testifying with appropriate scientific support. The existence and appropriateness of

106 As in Frye itself, there is a relevant scientific community (or communities since more than one field is involved) apart from the community of forensic scientists doing DNA analysis. It is these nonforensic communities to which a court should turn in deciding whether statistical probability calculations are, from a scientific standpoint, sufficiently well grounded under Frye (or FRE 702 as I think it should be read) to be admissible. Members of these scientific communities have testified on both sides of the population genetics issues.

107 I am assuming the lack of general acceptance I posit can be established. It is how I read the evidence of where the scientific consensus on these issues lies, but I do not claim to be an expert in these matters, nor to have read exhaustively in the relevant literatures. And as I indicated supra note 76, there is some recent research that suggests that the failure of these assumptions might not be as important in practice as it appears in theory. The question of whether a reasonable scientific consensus exists given recent research is one that courts appear to be well-equipped to determine in Frye or preliminary FRE 702 hearings. Given the rapid pace of research on DNA and the development of new techniques of analysis, no jurisdiction should be locked into holding as a matter of law that DNA evidence evaluated in a certain fashion is always or never admissible. There should be implicit in any law or precedent in this area the condition that the rule may always be reconsidered if it comes to be seen as scientifically unreasonable given new knowledge.

108 Where the matter relates to the weight that should be placed on a test that is scientifically well-grounded, a requirement for a scientific consensus on this issue may be too stringent a test for admissibility under FRE 702, but if there is a general consensus in the relevant community that a procedure for determining weight is not scientifically justified, testimony relying on the disapproved procedure should not be allowed. Ordinarily the relevant scientific community will not be the forensic science community but will be composed, as in Frye, of scientists who work in the fields in which the forensic science claims are rooted.

The arguments in the text and in this note are normative judgments and not an attempt to summarize what courts do. In fact with respect to DNA testing courts have come out both ways. As is clear from the cases cited supra at notes 101, 103, there are cases that have held that disputes about how to evaluate the implications of a DNA match go to weight and that parties can present competing evaluations to the jury; and there are cases, fewer in number, that have held that the admissibility of estimates of the weight to be placed on DNA evidence turns on the scientific acceptability of using the product rule and ignoring issues of population substructure.

Some might argue that Barefoot v. Estelle, 463 U.S. 880 (1983) (holding, in the face of an American Psychiatric Association amicus brief to the contrary, that a psychiatrist answering a hypothetical question could predict a defendant would be dangerous) is Supreme Court precedent favoring the admissibility of expert testimony on the weight to be accorded a DNA match even if the procedures for calculating that weight are based on assumptions and theories that are not generally accepted in the relevant scientific community, and that under Barefoot objections such as these to the scientific reliability of expert evaluations go only to weight and not admissibility. The strong version of the argument is that in Barefoot the expert evidence that
the claimed scientific support are matters that go to admissibility.\textsuperscript{109}

was offered (a prediction of dangerousness) was generally regarded as having no valid scientific foundation by scientists in the relevant field and that even in these circumstances the Supreme Court held that objections to the scientific foundation of the testimony went to weight rather than to admissibility.

While there are dicta in \textit{Barefoot} that would support this argument, the case properly read does not. First, this was a constitutional case. The Court was only deciding whether a Texas decision to admit the evidence violated a defendant's constitutional rights; \textit{id.} at 905. The Court was not determining the appropriate evidentiary standard for admitting such evidence. Second, the Court took great pains to suggest there was no consensus in the relevant scientific community that dangerousness could not be predicted in the hypothetical situation that the state's psychiatrist addressed, and the Court pointed out that the author of a study that the defendant's expert called "excellently done" had concluded, despite his initial expectations, that in some circumstances violence could be predicted. \textit{id.} at 899 n.7. Finally, the Court saw the situation in \textit{Barefoot} as one in which it would be anomalous not to allow psychiatrists to testify because even lay opinion on future dangerousness was to be respected, \textit{id.} at 897, and the issue of whether lay opinion on this issue was admissible was seen by the Court as foreclosed by its decision in \textit{Jurek v. Texas}, 428 U.S. 262 (1976), which held that the likelihood that a defendant would commit further crimes was a constitutionally acceptable criterion for imposing the death penalty.

\textsuperscript{109} It is, however, legitimate for a proponent of DNA evidence to argue that even if certain assumptions that underlie a probability calculation are questioned in the relevant scientific community, as a practical matter violations of these assumptions do little to affect the accuracy of the conclusions offered. The Court should determine the validity of such arguments, paying attention to whether they are accepted in, or accord with theories accepted in, the relevant scientific community. \textit{See, e.g.}, sources cited \textit{supra} note 76.