way, then this suggests a violation of the assumptions. It may be that the alleles were assumed to be independent and they are in fact highly correlated in the database being used. (There are many types of correlations. Many are irrelevant in a particular case. A correlation detected in looking at the distribution of LRs in the database is in a sense the only correlation that matters in the case at hand.) Or there may be an important degree of population substructuring. Indeed, a robust alternative to the usual process (robust at least for large databases) is to assume the average LR in the population is the average in the database and apply Bayes' theorem accordingly. For example, such an approach does not require assuming independence. The only concerns with this alternative are statistical: sampling variability, which is more of a problem for smaller databases, and whether the sample is random. I do not mean to minimize the latter concern: for example, population substructuring is still a problem.

EXPERTS AND ADVOCACY: BIASED TESTIMONY

Experts appear in court at the behest of the prosecution or defense. An attorney whose client has been charged in a violent crime in which DNA evidence is introduced wants to find scientists who will refute the evidence. An attorney for the prosecution wants to find scientists who will support the evidence. Sufficiently diligent searches will be successful. Opinions vary in every science. It is possible to find an expert with any given opinion who will testify [see Begley (1993) for examples]. There are various motivations for experts to testify. Some are laudable: science, common good, protecting individual rights. Some are not: money, notoriety, frequent flyer mileage. In any case, testimony of experts is biased. Judges and juries may know this and so discount expert testimony, but why should a legal system encourage testimony that ought to be discounted?

The searches mentioned above are eased by examining testimony in previous cases. As a result, the same experts tend to testify in case after case. In effect they become advocates, advocates for or against a technology. They—their persona, their thinking, their science—become objects of attack and to defend. Biases become more and more serious. Any semblance of objectivity disappears.

There must be a better way, one with less blatant bias. While this venue is hardly appropriate for recommending revisions in practices of jurisprudence, I will do so anyway: Expert testimony should be evaluated by expert witnesses supplied by the court. Court-appointed experts would be required to listen to the testimony of the prosecution and defense experts. Then they would express their reactions and opinions. They would not be paid but would be obliged to serve, with their service counting as jury duty in the municipality of their residence. If a municipality has no qualified experts, the court would import such from nearby municipalities and reimburse travel expenses. No system is bias-free, but this policy would rid the system of some extreme forms of bias.

Comment: Theory and Practice in DNA Fingerprinting

Richard Lempert

Throughout her useful paper on DNA identification, Professor Roeder properly attends to both theory and practice. Thus she acknowledges the theoretical soundness of certain criticisms that have been made of the standard paradigm used to evaluate DNA random match probabilities but argues that in practice these criticisms matter little. I am thinking here of the arguments that those cautioning against overweighing DNA evidence have made regarding the undeniable existence of population substructure and its potential implications for independence assumptions supporting the application of the product rule and for the use of convenience samples, such as data garnered from no more than a few local blood banks, to generate estimated allele frequencies for all Caucasians or African-Americans or Mexican-Americans living in the United States. Like Professor Roeder, I believe that these theoretically sound objections have, to date, been shown to be relatively unimportant in practice.

Richard Lempert is Francis A. Allen Collegiate Professor of Law and Professor of Sociology at the University of Michigan, Ann Arbor, Michigan 48109-1215. He is also acting chair of the Department of Sociology.

It is precisely because her attention to the implications of practice for theory are a strength of her paper, that I was disappointed by Professor Roeder's failure to attend to the full practical implications of certain issues she identifies. Let me give some examples.

REFERENCE POPULATIONS

First, in discussing the appropriate reference population, Professor Roeder, repeating an argument that Weir and Evett (1992) made, argues that if an innocent suspect is unrelated to the donor of an evidentiary sample, the suspect's ethnicity is irrelevant. In practice this is not true, for a suspect's ethnicity can matter when taken in conjunction with the composition of the population of potential suspects (henceforth "suspect population"). Assume that a crime occurs on an isolated Indian tribal reservation so that 99% of the potential perpetrators are members of the reservation tribe. DNA evidence is available, but the only data base from which to draw allele frequencies is a Caucasian one. Using this data base rather than a tribal data base is in theory inappropriate regardless of whether the suspect is Caucasian or Indian, but in practice one need only worry if the suspect were Indian. If the suspect is Caucasian, using estimated allele frequencies from the Caucasian data base rather than from a tribal one is almost certainly going to overestimate the probability of a match, which is to the suspect's advantage. Moreover, the Vermont case Professor Roeder uses to illustrate her point is an infelicitous example, for as Lewontin (1993) noted, the reference population issue is potentially substantial because the suspect, along with most of those in the suspect population, seem to have shared a similar Native American heritage.

Arguably Professor Roeder avoids my criticism of her treatment of the Vermont case and of the issue generally by her assumption that the suspect and culprit are unrelated and by her counterintuitive definition of "unrelated" in Section 4.2 of her article to include not belonging to the same subpopulation. However, these moves which save the argument at the level of theory destroy it in practice for one can seldom, if ever, be sure that suspect and culprit, assuming the suspect's innocence, are "unrelated" in the broad sense of not belonging to the same subpopulation.

RELATIVES

Second, Professor Roeder properly notes the implications of the possibility that the culprit is a suspect's relative for the usual statistics associated with DNA identifications. She goes on, however, to suggest that this is not a problem in practice because prosecutors can either ensure that all near relatives have a solid alibi or test the DNA of near relatives to show it does not match, and, if neither of these have been done, prosecutors will, she tells us, present the jury with statistics under the two competing hypotheses. It is unclear whether Professor Roeder means to suggest that these procedures are always followed when a suspect has close relatives living nearby or that they are only followed in those rare cases where a relative is also a suspect. If she means the former, I think she is wrong, for my impression—confirmed by a defense lawyer and a prosecutor, both of whom are experts on DNA cases—is that the presence of relatives is seldom factored into the statistics DNA experts present, nor do the police, once they have focused in on a suspect, routinely exclude all the suspect's relatives as possible perpetrators.

There are interesting statistical issues here as well as serious practical problems. Consider the simple situation to which Professor Roeder's discussion most directly pertains. Here the suspect population consists of two people, the defendant and another. Following Professor Roeder's method of culprit-based sampling (i.e., conditioning on the suspect's genotype, which is known to match the evidence genotype), the probability that a random draw from the suspect population will match the evidence DNA is the frequency of that genotype in a population data base that encompasses both the suspect and the stranger. If, however, the other is the defendant's brother, the probability of a matching genotype is far higher, as Professor Roeder has shown, because the chance that the brother's genotype will match the evidence DNA is not conditionally independent of the suspect's match.

Now consider a suspect population of a million, one of whom is the defendant's brother. If there is a random draw from the suspect population, the probability of a match with the evidence DNA is similar to what it was in the stranger case because the suspect's brother is very unlikely to be chosen at random. Thus where there are a large number of potential suspects the presence of a brother in the suspect population appears not to matter, but this argument assumes that the degree of a priori suspicion is the same for all possible suspects. A known suspect's brother is ordinarily a far more plausible alternative suspect than a random member of the suspect population because he is likely to share features (appearance, accent, gang membership, place of residence, etc.) that made his brother a suspect in the first instance. Thus DNA statistics should account for the presence of brothers and close relatives in the suspect population even when that population is large; but the appropriate degree of accounting is unclear.

Consider also a suspect population which, excluding the suspect, consists of 10 equally plausible sus-

pects: the suspect's brother and nine of his cousins. The conventional model suggests that a DNA match is more probative of the suspect's guilt in these circumstances than when the defendant's brother is the only alternative suspect, for the likelihood of matching the evidence DNA drawing at random from this population is less than the likelihood of a match where only the brother could be drawn. Yet surely the suspect's claim of innocence is more plausible when his brother and cousins might have left the evidence DNA than when only his brother might have left it. The key to this puzzle is the separation of likelihood ratios from prior probabilities. The prior probability of the suspect's guilt will be sufficiently less when his brother and nine cousins might be guilty than when his brother is the only plausible alternative suspect that the posterior probability that the evidence DNA was the suspect's will be less as well. The challenge is how to account for this in a system which only presents juries with likelihood ratios or their equivalents.

Giving jurors separate random match and relative match probabilities, an alternative I once endorsed (Lempert, 1991), is likely to leave a jury not knowing what to do with the data they are given; it may also impose on defendants a task they are ill-equipped to accomplish—showing a relative did it. A more defensible solution is to provide juries with information about what random match and relative match statistics taken together mean for the probability that there are at least n persons with matching DNA in the suspect population. This would, however, require a reasonable estimate of the size of the suspect population and the specification of those relatives who are in it. It would also risk what Thompson and Schumann (1987) call the prosecutor's fallacy, the tendency of jurors to treat evidence which, for example, limits the potential perpetrators of a crime to a defendant and one unknown other as not very probative because it seems to suggest that there is only a fifty percent chance that the defendant committed the crime.

Clearly more thought must be given to the issues relatives raise. Ultimately, a solution to this problem that is both practically viable and scientifically defensible may have to await either scientific advances that will allow DNA to be sequenced or procedural changes that will allow enough alleles to be typed so that when suspect and evidence DNA match, the probability that a relative's DNA might also match is minute.

ERROR

By far my most serious practical difficulty with Professor Roeder's argument is her failure, in discussing laboratory error, to state the obvious fact that *the incriminatory value of a DNA match can never be greater than the false positive error probability*. Given that false positive error rates with DNA analysis appear small, this would not be a serious problem except that, as Professor Roeder's discussion of uniqueness suggests, the random match probabilities DNA evidence yields are smaller than any plausible false positive rates by many orders of magnitude. Statisticians and geneticists involved in the controversy over DNA testing have understandably been fascinated by and mostly written on disputes regarding the statistical and genetic issues that DNA identifications raise, but laboratory error places the most serious limits on the evidentiary import of reported DNA matches. If justice is the mutual goal of those involved in the debates over DNA identifications—and I believe it is everyone's concern—the possibility of error must be honestly faced, and it must be incorporated into estimates of the incriminatory power of DNA matches.

Like Professor Roeder, I do not think the answer will be found in proficiency testing, for I doubt whether such testing will ever be extensive enough to generate reliable false positive probabilities. However, unlike Professor Roeder I think a rigorous program of blind proficiency testing is important. First, if a laboratory knows that any samples it analyzes may be a test of its skills, it will have a powerful incentive for care in all its work. Second, proficiency testing may allow the setting of bounds on likely false positive error rates. These will, no doubt, be many orders of magnitude higher than the random match probabilities associated with DNA evidence. However, since most causes of false positive error seem likely to be independent across laboratories, if casework DNA were routinely analyzed by two laboratories and both called matches, error rate bounded match probabilities would be quite low since they would be close to the product of the laboratories' individual error probabilities (Lempert, 1994). Unless laboratory error is uncomfortably common, these estimates, while not approaching the dazzling 1 in n billion estimates that are sometimes given for DNA random match probabilities, should be low enough to sustain convictions in all cases except those where the evidence, apart from the DNA match, strongly suggests that the defendant is innocent.

The problem of false exclusions also deserves attention. Acquitting guilty defendants who may rape again or even kill imposes potentially huge costs on society. Multiple tests or testing more alleles, while protecting the innocent and, ordinarily, further incriminating the guilty, will increase the chances of false exclusions since there will be more opportunities for an analyst to conclude that evidence and suspect DNA do not match. No one to date has con-

sidered what appropriate trade-offs between Type I and Type II error in this situation might be. Perhaps extant data and the kinds of modeling methods that Professor Roeder alludes to will allow us to begin empirically exploring this problem. Proficiency testing may also have much to offer. Since false negative error is likely to be more common than false positive error, proficiency tests might yield more reliable estimates of false negative error rates as well as some sense of the likely ratio between the two types of error.

THE NRC REPORT

Before concluding, I would like to comment on two criticisms Professor Roeder makes of the NRC report. First, she criticizes the report's concern for objective matching criteria. Part of her argument is that objective matching criteria will prevent subjective exclusions. This argument is a straw man. Prosecutors will not introduce DNA evidence when analysts tell them that they cannot conclude that the suspect and evidence DNA are identical even though formal matching criteria are met. The converse situation, where objective matching criteria are not met yet the analyst declares a match, has potential problems which justify the NRC's cautions. At a practical level, it is common for laboratory analysts to know something about the non-DNA evidence that implicates suspects. Thus, analysts' subjective judgments about the implications of near misses are likely to be swayed by what else they know about a case. Information about case quality should be denied analysts if there is any subjective element in declaring matches. Also, allowing subjective judgments makes match probabilities for a suspect's alleles higher in practice than objectively defined bin frequencies would suggest. An advantage of Bayesian approaches that obviate the need to declare matches is that they eliminate some of these problems.

Professor Roeder also faults the NRC report for its treatment of Bayesian approaches. I agree in part with her criticism. At one point in the report it states that Bayesian approaches are not considered because they are not commonly employed in the presentation of DNA evidence. Given the practical concerns of the NRC committee and that committee's makeup, this decision was entirely appropriate. At another point, however, the report appears to reject Bayesian approaches. Here the committee had no adequate justification, and Professor Roeder's criticism is fair. Indeed, like Professor Roeder I think that Bayesian approaches have much to offer in this area. However, if they are to be used, it is not enough to consider likelihood ratios. As the "relative" issue illustrates, attention should be paid to what constitutes prior probabilities in DNA cases and to how jurors may be aided in assessing the evidentiary implications of likelihood ratios given the different prior probabilities they hold (cf. Kaye, 1994).

CONCLUSION

With some notable exceptions (e.g., Evett, 1992b) there has been a partisanship apparent in almost all scientific writing on DNA evidence. Scientists who have cautioned about overweighing DNA evidence have struggled against rather than fully acknowledged the evidence suggesting the robustness of product rule procedures and convenience sample data bases to the theoretical threat posed by population substructure. Scientists promoting DNA identification have tended to downplay the importance of issues like error probabilities and the implications of relatives in the suspect population or to ignore them entirely. Moreover, there is a kind of passion to each side, which sometimes seems, however politely, to amount to questioning the bona fides of the other. Yet I think scientists on both sides have acted in good faith, and in this instance the importation of legal adversariness into the scientific world has spurred both valuable research and practical improvements in the way DNA evidence is analyzed and presented.

Professor Roeder's article places her closer to those who have been advocates for DNA than to those who have urged caution. Yet she is explicit in her recognition of the problems posed by relatives and error, even if she does not dig as deeply into these issues as I would have her do. Comments, by their nature, tend to focus on points of difference. Thus let me conclude by saying that overall I think Professor Roeder's clear exposition of statistically related issues surrounding DNA identification is a helpful, interesting and fair treatment.