the "RST" estimator that Rothman, Sing and Templeton (1974) derived from a Dirichlet model. The moment estimator was essentially unbiased for their parameter values whereas the RST estimator had about a 50% bias. The RST estimator had a standard deviation about half that of the moment estimator. Jiang and Cockerham concluded that the Dirichlet model performed poorly for the genetic drift process, and were concerned that the model may not be broadly useful.

Notwithstanding these comments, the paper by Roeder is a welcome addition to the literature. It illustrates the role statisticians have to play in addressing societal issues.

Rejoinder

Kathryn Roeder

I would like to thank the discussants for their lively remarks, even those wide of the mark. Because of the subject of my review, I am not surprised by some of the emotional arguments put forth, although they seem out of place in *Statistical Science*. As Professor Lempert comments, "there is a kind of passion to each side, which sometimes seems, however politely, to amount to questioning the bona fides of the other." Before discussing the commentators' remarks in detail, I will outline their points.

The discussants broach several interesting issues that are far afield from the points covered in my review. The statistical issues in human population genetics, the core of my review, have been the focus of controversy in the courts and the scientific literature for the last few years. Professors Berry, Lempert and Weir agree with me that the criticisms leveled at the standard paradigm for estimating DNA profile probabilities, while sometimes sound in theory, have a negligible impact on the calculations in practice. Professors Balding, Donnelly and Nichols (BDN) and Professor Lewontin continue to question some population genetic assumptions upon which the probability calculations are based. Professor Sudbury stands alone in questioning the need for the paradigm. The adequacy of the genetic model and the importance of the choice of reference population are elaborated in Sections 1 and 4 of my rejoinder.

Several commentators raise concerns about laboratory error, something I did not discuss in depth in my original paper. They worry that samples will be mixed up in the laboratory, resulting in the suspect's sample being compared with itself, rather than with the crime sample. Another concern they raise is cross-contamination, which could also lead to an erroneous match. Professor Lewontin says that the danger of this is greater when a molecular technique known as PCR is used. I think that the danger of error depends more strongly on laboratory protocol than on the molecular technique.

From his comments, it seems that Professor Lewontin is unfamiliar with the protocol and methodology generally used by forensic scientists. He asserts that crime scene samples, being of limited quantity, are amplified using PCR. In fact, PCR is generally not used for the purpose he describes, and the genetic evidence presented at trial is usually not the product of PCR amplification. The major forensic testing laboratories (FBI, Lifecodes and Cellmark) do not regularly use PCR now, let alone in the past [Ivan Balazs, Director of Research at Lifecodes, and Bruce Budowle, Director of Research at the FBI Laboratory (Balazs and Budowle, 1993)]. Although PCR is sometimes used for an initial screening, for the bulk of cases forensic testing laboratories ultimately use a less sensitive method called RFLP typing via Southern blotting (NRC, 1992). Perhaps Lewontin's remarks are aimed at what he envisions for the future. Indeed RFLP analysis will eventually be replaced by some amplification process because results for the latter can be obtained almost immediately, whereas results for any RFLP analysis require four to six weeks or more.

Professors Thompson, Lempert and Berry believe that the average probability of a laboratory error should place a lower bound on the probability of a match. I disagree. A case-specific, posterior probability of a laboratory error is the appropriate calculation. Such a calculation, if admissible in court, should be presented separately from the probability of a match. Relying on the NRC report, Professor Thompson argues that the probability of a laboratory error should be estimated using proficiency testing. From the statistical perspective, it is clear that proficiency testing is not an efficient means of estimating a small probability. In Section 5, I discuss laboratory error in general, including methods of estimating the probability of error.

BDN voice concern about likelihood ratio statistics that calculate the probability of a match between un-

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science*. www.jstor.org

related individuals, claiming they should account for the probability of a match between relatives. No one would argue with the fact that relatives are considerably more likely to have matching DNA than unrelated individuals. The disagreement lies in how this information is to be presented in court. Lempert suggests that this issue is serious enough to postpone the use of DNA fingerprinting until there is virtually no chance that relatives would match. My view is that the importance of this issue is exaggerated because the courts already understand the issue of relatives. Whenever it is appropriate to consider relatives, two approaches are available: supplementary probability calculations may be presented, or further DNA testing can be conducted (see Sections 2 and 3). When specific relatives are under suspicion, the best solution is to obtain their DNA fingerprints.

The commentators, in general, wish to discuss how evidence should be presented in court. When addressing legal issues, I will restrict myself to the U.S. judicial system (see Sections 2 and 7). For this topic, it is important to distinguish between the realms of science and the court. In the courtroom, defense lawyers (and others providing defense counsel) can and should use all appropriate means to debunk the evidence against their client. Alternatively, it is the job of the prosecution to present as compelling a case against the defendant as possible, again using all appropriate evidence. The role of scientists is to evaluate the available data to answer questions of interest to legal scholars, and to suggest experiments and other forms of study when relevant data are unavailable. Several of the discussants fail to distinguish between the legal and scientific realms, and between the responsibilities of the defense and prosecution. These distinctions are critical, as we shall see in the sections involving interpretation of the evidence.

Finally, one might ask why this topic engenders so much controversy? Perhaps it is because the probability of a match can be so minute. The strength of this evidence stands in sharp contrast to much of the information jurors assimilate to make their decision. Yet similar kinds of evidence are routinely presented in court, including dermal fingerprints, hair matching, eyewitness identification and ballistic evidence. As I stated above, I think that it is the responsibility of the statistician to provide solid estimates of the relevant probabilities and to test the assumptions of the model. It is the province of legal scholars to determine how conservative these probabilities should be. Once a probability enters into court, it is the court's responsibility to ensure that this information is integrated into the case in a balanced way. In this regard, it is important to remember that cases will rarely, if ever, be based entirely on DNA evidence. In a typical case, DNA evidence only provides circumstantial evidence that potentially places the defendant at the scene of the crime; in this instance, DNA evidence, by itself, cannot lead to a conviction.

1. POPULATION GENETICS

I am pleased that Professor Weir agrees with me on the majority of the population genetic issues relevant to forensic inference. For some time now my colleagues and I have been tackling many of the same research problems as Weir, using different techniques, but nearly always reaching the same conclusions. I thank him for referring me to some of the more specialized results from the population genetic literature. I agree with him that the Dirichlet model can only provide an approximation to reality. Nevertheless, I think it offers a useful first-order result concerning the amount of heterogeneity in the population.

The question of independence, once a hot area of controversy, appears to be settled. Only BDN bring up this issue, claiming there is not enough data to obtain sufficient power in tests of independence. The record shows the opposite: some databases are becoming so large that even the most negligible violations of independence will be detected. As Berry notes, it is the impact of the violation, not the presence, that matters.

In his commentary, Professor Lewontin states that "all parties agree that the differentiation among [major ethnic groups] is as large, if not larger than, the differences among tribes and national groups [within a major ethnic group]." I am pleased that he has changed his mind. Unfortunately, Lewontin and Hartl (1991) previously claimed "there is, on average, one-third more genetic variation among Irish, Spanish, Italians, Slavs, Swedes and other subpopulations, than there is, on average, between Europeans, Asians, Africans, Amerindians, and Oceanians." More importantly, the NRC report echoes Lewontin and Hartl's remark, arguing that it is well known that there is more variation between subpopulations than among the major ethnic groups. In fact, this claim was the motivation for the NRC's ceiling principle (Devlin, Risch and Roeder, 1993a, b).

Differentiation among tribes and national groups is trivial for some major ethnic groups (Caucasians, African Americans), quite small for other groups (Hispanics), but more important for the tribes of Amerindians (Chakraborty 1993; Morton, Collins and Balazs, 1993; Roeder et al., 1993; Weir, 1994). With respect to the heterogeneity at the level of major ethnic groups, Lewontin has reversed opinions again. He now claims, based on his 1972 research article, that there are large differences among the major ethnic groups. However, in 1972, Lewontin recognized there were not large genetic differences among the major ethnic groups: "It is clear that our

perception of relatively large differences between human races and subgroups [is] a biased perception and that, based on randomly chosen genetic differences, human races and populations are remarkably similar to each other, with the largest part by far of human variation being accounted for by the differences between individuals." This is precisely the paradigm that forensic laboratories rely on to estimate DNA profile probabilities.

Laboratories that perform forensic testing maintain separate databases for the major ethnic groups, African Americans, Asian Americans, Caucasians and Hispanics, as well as other databases. The jury is usually provided with the profile probabilities calculated based on several of these databases. Consequently, jurors are able to view the extent of the variability in the DNA profile probabilities, which are generally similar, contrary to Lewontin's claim that they are widely different (see, e.g., Chakraborty and Kidd, 1991, Figure 1).

Also contrary to Lewontin's argument, no one advocates weighting the probabilities by the proportion of each ethnic group residing in the proximity of the crime scene because no objective definition of what constitutes the proximity of the crime scene can be offered in most cases. Moreover, since emphasis is placed on the least incriminating probability, the method of presentation currently in use is conservative.

While Lewontin recognizes the fact that there is less heterogeneity among subpopulations than among the major ethnic groups, BDN believe that there are large differences among subpopulations that we have not yet measured. Whenever the heterogeneity for subpopulations of major ethnic groups is quantified, be it for standard genetic markers or for VNTR markers, the estimated amount of heterogeneity is not large. This conclusion is not based on a few scattered studies (or a sample of 56 individuals from Glasgow); there are numerous studies of population heterogeneity. The most recent studies, using VNTR loci, place $\widehat{\theta}_s$ approximately 0.002 for Caucasian and African Americans; the values are larger for Amerindians ($\widehat{\theta}_s \approx 0.02$) (Chakraborty, 1993; Morton, Collins and Balazs, 1993; Roeder et al., 1993; Weir, 1993). These values are in stark contrast with the value initially advocated by Nichols and Balding (1991), $\theta_s = 0.05$, which has gained some acceptance in the U.K. court system, at least for some cases.

In fact, $\theta_s = 0.05$ is larger than the value Roeder et al. (1993) estimate from artificial mixtures of extremely heterogeneous populations composed of groups from all over the world. BDN fault this analysis, saying that the methods are sensitive to modeling assumptions. While this may be a fair criticism, it is contradicted by the fact that nearly all estimates of heterogeneity are approximately equal to the values obtained in our study, regardless of the type of genetic markers evaluated or the modeling assumptions imposed. For example, Weir (1993) does not use the Dirichlet assumption or the same databases, yet he obtains results nearly identical to those in Roeder et al. (1993). In addition, these small values of population heterogeneity are supported by relevant information as diverse as population demographics and marriage statistics (Lieberman and Waters, 1985; Alba and Golden, 1986; Alba, 1987; Morton, 1992).

BDN also advance the notion that heterogeneity may be greater for divisions of populations even finer than subpopulations, such as villages. The human genetics literature suggests otherwise. Even quite remote villages in Brazil (Smouse, Spielman and Park, 1982) and the tribes of India (Murty et al., 1993) show only small amounts of heterogeneity at this level. The same is true for VNTR markers: estimates of heterogeneity between locations in the United States for the same ethnic group are miniscule (Weir, 1993; Chakraborty, 1993).

To support their positions, BDN and Thompson cite Krane et al. (1992), a paper that I ignored in my review. BDN acknowledge "criticisms of the study," but say the study suggests "a level of differentiation which is sufficient to cause concern about current forensic practice." That is a strong statement to make on the basis of statistical artifacts. One of the alleged results of the Krane paper (Hartl and Lewontin, 1993) was a demonstration of "large" differences in profile probabilities among Finns and Italians, large enough to justify the need for very conservative calculations. They claimed that, 77% of the time, a Finnish profile was more common in a Finnish (or cognate) database than it was in an Italian (or noncognate) database, and vice versa.

Unfortunately, Krane et al. overlooked sampling theory. Both the Finnish and Italian samples were extremely small, and the individual whose profile was to be estimated was left in the database. Naturally this induces a large bias toward larger profile probabilities in the cognate database [see Devlin, Risch and Roeder (1993b), for discussion]. Budowle, Monson and Giusti (1994) quantified the bias by using resampling methods and Krane et al.'s much larger Caucasian database ($N = 1353$). From the Caucasian database, they created two small databases of similar structure to the Finnish and Italian databases and then quantified profile probabilities. Naively, 50% of the profiles are expected to be more common in the cognate database; however, the bias caused the observed value for 1000 experiments to be 73.3%, with a standard deviation of 4.4%. Thus the reported evidence for heterogeneity is artifactual.

This is not the only statistical flaw in the Krane

et al. (1992) paper. For the VNTR locus HRAS1 and their Caucasian population, they reported a highly significant intraclass correlation of 0.26. This result was interpreted as indicating substantial subpopulation heterogeneity. Further analysis of the data revealed that the large correlation was due solely to six outlier pairs of observations, subsequently confirmed by the laboratory to be data transcription errors (Devlin, Krontiris and Risch, 1993). After deleting these six improper pairs, the correlation plummeted to 0.03. By studying the HRAS1 allele distributions for various European populations, Devlin, Krontiris and Risch estimated $\hat{\theta}_s = 0.003$, indicating almost no heterogeneity among these populations.

Professor Sudbury presents a completely different view than the other commentators. He advocates ignoring the results of population genetics altogether. Given the brouhaha over population genetics, his approach is a refreshing respite. Provided the suspect's DNA profile matches the culprit, but not any other sample in the database (of size m), Sudbury asserts that an appropriate p-value is $2e^{-1}/m(m+1)$. This formula has the appeal of simplicity, which is indeed valued by the courts. It is distinctly counterintuitive, however. As a p-value, this measure of the strength of the evidence depends more strongly on the size of the database than on the particulars of the crime sample, such as the number of loci typed. For instance, with $n = 100$, there probably will not be any matches in the database if two or three VNTR loci are used. The resulting p-value is approximately 0.0001. Now suppose 10 loci are tested and the suspect still matches. Ten loci might be sufficient to identify the genome uniquely (except for the occasional brother), yet the evidence conveyed by the p-value is no stronger.

To implement this formulation, the first question we might ask is, What is m? Is it the number of people in all possible databases, assuming there were no matches in any of them? Or should we only consider those people in the ethnic group (or even the subpopulation) of potential perpetrators? The answer to this question can only be found by ascertaining the amount of heterogeneity among and within populations. Thus we must rely on population genetics.

2. INTERPRETING LIKELIHOOD RATIOS

Before launching into a discussion of likelihood ratios and Bayes' theorem, I present an excerpt of the advice read to jurors in the state of California: "[B]efore an inference essential to establish guilt may be found to have proved beyond a reasonable doubt, each fact or circumstance upon which such inference necessarily rests must be proved beyond a reasonable doubt. Also, if the circumstantial evidence is susceptible of two reasonable interpretations, one of which points to the defendant's guilt and the other to [his] [her] innocence, you must adopt that interpretation which points to the defendant's innocence." Notice that each fact or circumstance of the case, not just the DNA evidence, must be proved beyond reasonable doubt. Furthermore, the case, it states, is to be viewed as a whole, dependant on its weakest link. This implies that no portion of the case is to overwhelm the rest. This argues against combining evidence using Bayes' theorem. Apparently, the probabilistic concept of coherency, which we statisticians strive for, is not the objective of the judicial system.

I agree with BDN that hypothesis testing is not the way to present the DNA evidence. Likelihood ratios, properly interpreted, contain all of the information that an expert witness can convey to the jury. The likelihood ratio quantifies the circumstantial evidence that attempts to place the defendant at the scene of the crime.

To make clear a subtle, but frequently misinterpreted, point about how genetic evidence should be interpreted, consider a somewhat different type of evidence, dermal fingerprints. Suppose a murder has occurred, and in the course of their investigation the police discover that the neighbor's fingerprints are on a water glass in the victim's home, where the crime occurred. This evidence, along with other evidence potentially incriminating the neighbor, is presented at trial. The prosecuting attorney, in presenting the evidence, tells the jurors that it is infinitely more likely to have obtained this data if the defendant's finger touched this glass (H_1 in my notation) than if the fingerprint was from any other person (H_0). Can the jurors safely conclude from this fact that the defendant is guilty? Of course not. Although the prosecuting attorney's likelihood ratio is correct, it is by no means proof of guilt. Any defense attorney is bound to provide numerous harmless reasons why the neighbor's figerprint was on the glass and will undoubtdly point out that this evidence, at best, places a water glass bearing the neighbor's print at the crime scene.

Now suppose, instead of a neighbor, it is the fingerprint of an individual who lives in another city that is found on the water glass. In addition, suppose there is no obvious reason for this defendant's presence in the victim's home. Will the jurors conclude from this fact that the defendant is guilty? Not necessarily. Like the previous scenario, this evidence only places a water glass bearing the defendant's print at the crime scene.

The analogy to DNA profiles should be clear. The likelihood ratio for DNA profile evidence is a means of providing the court with an estimate of how much more likely it would be to observe the evidence if it is from the defendant than if it is from some other,

appropriately chosen, source. Like dermal fingerprints, DNA profiles can only potentially place the defendant at the crime scene; unlike dermal fingerprints, the partial DNA fingerprints currently in use are not unique. DNA fingerprints, as currently defined, can only provide circumstantial evidence that places the defendant at the crime scene, while dermal fingerprints provide direct evidence of this fact. Again, quoting from the advice to jurors, "[A] finding of guilt as to any crime may not be based on circumstantial evidence unless the proved circumstances are not only consistent with the theory that the defendant is guilty of the crime, but cannot be reconciled with any other rational conclusion."

Some commentators write statements that suggest the disposition of the case rests on the interpretation of the DNA profile. While it may be true that the strongest evidence in some cases is the DNA, this would not be the entire basis of any case. As noted above, DNA evidence only offers circumstantial evidence that places the defendant at the crime scene. Such a case would be weak regardless of the size of the likelihood ratio. Many cases more closely resemble three cases recently reviewed by the California Appellate court.[1] For each case, Justice Chen ruled that it was error to admit the DNA evidence at trial (because of the population genetics "controversy"), yet he also ruled that the error was harmless in each case because the other evidence was overwhelming. In many cases in which the DNA evidence was not admitted, convictions have been obtained; conversely, in some cases in which DNA evidence was admitted, a conviction has not been obtained. In at least one case, the defendant was convicted (of rape) even though his DNA did not match the crime sample.

BDN present an interesting formula to assess the probability of guilt in a coherent way by combining genetic and nongenetic evidence. The method directly relates the genetic evidence to the probability of guilt. This method requires a specific input of prior guilt for each and every potential culprit. The resulting odds of innocence are a weighted sum of prior odds and likelihood ratios for various hypotheses (culprit is a brother, culprit is an uncle, culprit is from the same subpopulation and so forth).

In Section 4, I gave specific formulas from which the likelihood ratio could be calculated for different hypotheses and directed the readers to a more in-depth discussion of this topic. BDN ask how would a juror process information concerning the competing null hypotheses without applying their (1). I believe BDN are incorrect, however, when they make a direct connection between guilt and the evidentiary sample. To apply Bayes' rule correctly in this problem, a few more logical steps are required to connect the event "guilty" to the event "suspect's DNA found at crime scene."

If there is reason to suspect a close relative (i.e., brother, father or half-brother), (1) will be dominated by the term(s) involving this close relative(s). It is also worth noting that, on close inspection, (1) is almost identical to the model considered by Evett (1992b) for instances in which one of several close relatives is possibly the perpetrator. However, because of the difficulty of assigning priors, Evett (Chief Statistician, Home Office Forensic Science Services) rejected the model in favor of direct testing when close relatives are potential suspects. I agree with Evett that close relatives, when under suspicion, should be tested. Prosecutors will routinely test close relatives under suspicion because it strengthens their case. Professor Lempert worries that they will not. In such situations, the defense can capitalize on the prosecution's foolhardy behavior.

BDN and Lempert also worry about other relatives, even relatives on an evolutionary time scale. From my equation (19), it is obvious that the correlation plummets with increasing degree of relation, and the subpopulation calculation (18) yields an excellent approximation for anyone more distant than a first cousin.

I also disagree with BDN about the relevant use of DNA evidence in the legal setting. The process by which jurors (or justices) reach a decision is complex, and formal probability arguments undoubtedly never enter the process. In fact, for criminal cases in the United States, formal probability arguments for combining evidence to determine the posterior probability of guilt, with priors provided by the expert witness, are not admissible because they are thought to usurp the jurors' role. BDN pay lip service to this fact, then ignore it and discuss methods that require sophisticated prior input that no juror (including me) could manage. It is difficult to imagine how to evaluate (1) when default priors are not permitted.

Experiments have been conducted to determine how jurors process probabilistic evidence when other sources of evidence are available. These studies indicate a tendency to underweight the statistical evidence (Kaye and Koehler, 1991), contrary to the claims of Lempert and BDN. One study estimated the susceptibility of mock jurors to equate $\Pr(\mathcal{G} \mid H_0)$ with $\Pr(H_0 \mid \mathcal{G})$ (the prosecutor's fallacy) at only 5%. Of course more experiments of this sort should be performed because, as Kaye and Koehler acknowledge, each of these studies possessed some shortcomings. In particular, none of these studies measured the effect of extremely small probabilities on jurors' decisions. Nevertheless, these results suggest that Berry's and Lempert's concern that the jury is unable to process the data is exaggerated. I am, of course,

[1] 1992 WL 184530 (CA App. 1 Dist.).

in agreement with Berry, that the evidence must be explained properly to the jury if we expect them to process it correctly. Finally, I want to emphasize that these mock jurors were forced to evaluate their evidence in a probabilistic manner. These experiments shed little light on the actual mechanism by which a jury reaches a conclusion. It seems much more likely that jurors use gestalt rather than formal probability arguments to arrive at their decisions.

BDN also note, based on a hypothetical scenario, that there can be a substantial difference between a likelihood ratio of 100,000 and 10,000,000. They reason that, if the prior is 1/10,000, the former likelihood ratio could be grounds for an acquittal whereas the latter could be grounds for a conviction. It is a dubious presumption on their part, however, to suggest that jurors process information in this way.

As an aside, the quote of mine by BDN on this topic reminds me of the famous remark "Please don't misquote me out of context." BDN say that I argue that "errors" of one to two orders of magnitude are not meaningful as long as the likelihood ratio is big enough because it will have little impact on the jurors' decision. While it seems likely that the argument is correct, that is not what I said. What I said was that, for extremely large likelihood ratios, the variability among the major ethnic groups was approximately one to two orders of magnitude, so that even this variability does not have a major impact on the likelihood ratio—not the decision process of the jurors. In general, the jurors are told that the likelihood ratio varies by one to two orders of magnitude, depending on the ethnic group, and are provided with the numbers themselves (as well as with other measures of variability). In addition, the defense will undoubtedly emphasize the variability as much as possible. Therefore, the jurors will have ample opportunity to understand that the likelihood ratio is only an estimate that depends somewhat on the population chosen.

BDN also imply that the likelihood ratios presented in court would be identical under two distinctly different scenarios: a DNA profile matching the evidentiary sample is obtained from the defendant only with probable cause; or a match between evidentiary and defendant samples is found by a search of the DNA database. As Berry notes, the NRC panel recommended such data not be used at all. Certainly these two scenarios are not handled identically in the United States.

I enjoyed Professor Berry's discussion of suspects identified by searching through a database even though much of it was impractical and/or inadmissible, so far. He presents two scenarios that result in a circumstantial case against his hypothetical Mr. Pritchard. In his example, two pieces of data are used to build a case: the first for hypothesis generation, the second for hypothesis confirmation. His hypothetical case is similar to *Minnesota v. Perez*.[2] Two loci were used to identify the defendant from a database; then six other VNTR loci were used for confirmation. Only the results of the confirmatory tests were presented in court. As Berry notes, a Bayesian would treat both portions of data symmetrically.

Now consider the situation in which all of the loci were used to identify a suspect from a database. Is this data less valuable than data obtained in a hypothesis-confirmatory mode? Of course the prior probability of H_1 is much smaller than if the data were obtained in a confirmatory mode, but what of the likelihood ratio? As Berry indicates, the inferential aspects of database searches are problematic for frequentists. This is not the appropriate forum for a Bayesian–frequentist debate. I admit, however, that I am uneasy about ignoring the sampling mechanism by which the data are obtained, especially in light of the restrictions placed on Bayesian inference in the legal setting. Theoreticians are free to ignore these complex issues if they take a Bayesian perspective. Hence Berry can proceed to combine the likelihood ratio with nongenetic evidence, ultimately obtaining the posterior probability of H_1. He attempts to avoid the limits of the "island problem" by incorporating more nongenetic information into the probability model; for instance, he allows for an open population. The island problem assumes every person in the population (island) is equally likely to be the culprit (Dawid, 1994). Because human populations in industrial societies are not closed, even the population size may be unknowable. Hence the solution to the island problem is impossible to obtain in practice. In my opinion, further study is required, and this is why I placed the topic of searching DNA databases in the "Open issues" section of my review. It is a topic worthy of serious consideration by statisticians.

3. RELATIVES: A LEGAL PERSPECTIVE

As I noted in my conclusions, ad hoc methods such as the ceiling principle can obscure the need for corrections, such as those described in Section 4.3, when the suspect and the culprit are assumed to be close relatives. Professor Lempert is correct in stating that the prosecution does not need to exclude all close-relatives as possible donors of the evidentiary sample, nor do they necessarily present the jury with the likelihood calculations tailored to the close relative hypothesis. The defense, however, can (and usually does) raise this issue during cross-examination of the prosecution's expert witness. If the defense can convince the jury that the brother question is relevant

[2] *Minnesota v. Perez*, "SIP" 92003535 (County of Hepepin, 4th Judicial Dist.).

and the prosecution has failed to exclude brothers, then this creates reasonable doubt potentially sufficient for exoneration. This is the very essence of the adversarial system. The specifics of the case will dictate which hypotheses are relevant, and the defense and prosecution will respond accordingly.

Lempert devotes a substantial portion of his commentary discussing relatives, concluding that the issue must be given more thought and the solution to this problem may have to await "procedural changes that will allow enough alleles to be typed, so that when suspect and DNA evidence match, the probability that a relative's DNA might also match is minute." Indeed, the subject deserves serious thought, although I disagree with the commentators who seem to argue that the "problem" of relatives is something novel to DNA typing. To the contrary, the courts have been dealing with the problem of relatives since the inception of the rules of law. As Lempert emphasizes, it is sometimes true that relatives live in the vicinity of the crime and may have had access to the crime scene. Relatives also look similar to the defendant, they have dermal fingerprints similar to the defendant (and therefore their partial prints might be indistinguishable) and their motives can be similar. Finally, since the evidence on less polymorphic genetic markers has been admitted as evidence for the past 50 years, it is noteworthy that the probabilities associated with those markers are affected in exactly the same manner as described previously. All of these "problems" concerning relatives have not brought the criminal justice system to its knees. Even without DNA evidence, trials proceed, defense lawyers raise the issue of relatives whenever possible and prosecution attorneys rebut these arguments whenever they can. In the process, defendants are convicted or exonerated by juries who understand the concept of relatives—they have them.

4. WHICH DATABASE?

The question "Which database?" is an interesting one. In my article, I tried to clarify a number of thorny issues. There can be no doubt that it is the population of potential perpetrators, not the population of the suspect, that is relevant for a particular case. It is sometimes difficult, however, to delineate the population of potential perpetrators. In some of these cases, it is convenient (and generally conservative) to use the ethnic group of the suspect, as well as several other ethnic groups that reside in the area of the crime scene. Furthermore, for some cases, there may be good reason to believe that the suspect and evidentiary sample are drawn from the same subpopulation. This was one of the three scenarios presented in Section 4 of my article. Lempert and Lewontin argue that corrections should be made to account for subpopulations. Clearly I agree that such cases exist, but the decision on whether a correction is necessary should be made on a case-by-case basis. In addition, subpopulation corrections are largely irrelevant for some major ethnic groups (Caucasians, African Americans) as there appears to be little variability among the subpopulations.

As a specific example of the issues created by subpopulation heterogeneity, Lewontin and Lempert return to *Vermont v. Passino*[3] (the Franklin County, Vermont, case) and argue with my assertion that the court's ruling was incorrect. In that case, the court decided that the DNA evidence was inadmissible because the FBI lacked a database composed of persons of the same mixed Abenaki, Italian and French ancestry as the suspect. [Mr. Passino's mother testified that he was $\frac{1}{2}$ Italian, $\frac{3}{8}$ Amerindian and $\frac{1}{8}$ French. No claim was made that his Indian ancestors were Abenaki (Hughes, 1993).] As Lewontin notes, the prosecution is required to provide a database representative of the pool of likely perpetrators, not the suspect. This database may or may not be of the same heritage as the suspect. In the particular case of *Vermont v. Passino*, it is beyond credulity to assert or require that the likely perpetrators were all of the same mixed Amerindian, Italian and French ancestry as the suspect. Therefore the judge's reasoning was fallacious. Whether or not the evidence should have been admissible, based on the argument that no Abenaki database existed, is a different question. Lewontin and Lempert fail to differentiate between the two.

In his opinion and order denying the introduction of DNA evidence in *Vermont v. Passino*, Judge Kilburn felt he might have come to a different conclusion if more information had been available to him. He states that the court must reach a different decision in this case than Judge Billings reached in *United States v. Jacobetz* partly because the court did not have sufficient access to the opinions of experts.

Indeed, no database had been collected for the Abenaki population at the time of this trial. If the defense argued persuasively that the culprit was likely to have been from the Abenaki population and that the Abenaki was a genetically differentiated group, then the prosecution would be compelled either to produce an appropriate reference population or to make adjustments to profile probabilities calculated based upon some other reference population, to account for alleles that may be more common in the local population.

The necessary adjustments are neither difficult nor of extremely large effect. To examine the is-

[3] *Vermont v. Passino*, 185-1-90 (Dis. Ct. Franklin County, VT, 1991).

sue of adjustments for population heterogeneity, the profile probabilities obtained from an extremely heterogeneous collection of 13 populations (Australian Aborigines, Navajos, Chinese and so forth) were compared to profile probabilities calculated based on the Caucasian population (Roeder et al., 1993). Because the actual profiles were of Caucasian derivation, the profiles probabilities were generally larger when calculated in the Caucasian database; however, the probabilities obtained using the mixed database, which ranged from 10^{-5} to 10^{-12}, were usually within one to one and a half orders of magnitude of the Caucasian profile probability. Using $\theta_s = 0.036$, the subpopulation correction given in my (18) provided an adequate correction, even in this extreme situation. Incidentally, 0.036 was the median of the posterior distribution of θ_s obtained for this artificially created mixed population. Thus, this result contradicts BDN's comment that evaluating (18) at the median will not be sufficient: in fact, we found that using the median induces a more conservative correction than integrating over the entire posterior distribution.

Because *Vermont v. Passino* is used by both Lewontin and Lempert as an ideal example of a "subpopulation" case, I conducted some modest research on the Abenakis. Based on the *Atlas of the North American Indian* (Waldman, 1985, pages 196–197), the Abenaki are not a federally recognized Amerindian tribe. The Abenakis are an organized group, however, and they continue to seek recognition. It is worth noting that an Aboriginal pedigree is not required to become a member of the Abenaki organization. In *Vermont v. Raleigh Elliott et al.*, the Supreme Court of Vermont heard a case concerning the lands occupied by the Abenakis. In their ruling, the Court concluded that all Aboriginal rights had been extinguished because there was a lack of tight historical ties (and presumably genetic ties) to the original tribe. In fact, there has been intermarriage between the Abenakis and Caucasians since the eighteenth century, which continues up to today. Hence it appears that the existent Abenakis are more accurately described by their distinct cultural affinities than by their genetic differentiation. It is inappropriate to equate cultural units such as the Abenakis with genetically differentiated tribes such as the Navajos.

5. LABORATORY ERROR

The topic of laboratory error garnered much of the commentary. One could take this as ample validation that its placement in the "Open issues" section was appropriate; or that the sources of substantial controversy in population genetics have been exhausted now that the facts and figures have been martialed. My comment on the mistaken emphasis on proficiency testing was much ballyhooed and generally misunderstood.

The source of the confusion on this subject can be traced to the NRC report. As Professor Thompson says, "According to the NRC report, accurate estimates of error rate require proficiency tests that are externally administered, are blind and are based on samples that are truly representative of case materials." To the contrary, statisticians know that direct efforts to measure the probability of a rare event through proficiency testing are inefficient and inherently inaccurate. They have experience with this type of problem (e.g., estimating the probability of a shuttle disaster) and can do better. It is unfortunate that the NRC chose not to place at least one statistician on a committee dealing with statistical issues. Typically, multiple sources of related information are available from which the probability of a rare event can be estimated, of which blind testing is only one. Other information can be obtained from paternity testing, from failure rates for related endeavors and the like. Combining this information yields a much more accurate measure of the prior probability of laboratory error.

A plentiful source of data is available from which laboratory error rates could be assessed. Data obtained from paternity suits are nearly ideal for estimating handling errors. On a single gel four samples of DNA are compared: the child, the mother, the putative father and a combination of the child and father. Mix-ups would be immediately apparent and should be available from the records of the company. Tens of thousands of such paternity tests have been performed, which could provide a rich source of data for this purpose. In addition, paternity tests that are disputed are frequently retested by another laboratory; here again, mix-ups are readily apparent. The drawback to these data, as noted by Thompson, is that the DNA samples are from fresh blood, which is not true for all crime scene samples.

The other key point concerning laboratory error, which was missed by the NRC report, is that the probability of an error varies on a case-by-case basis. For instance, if the chain of custody is broken, then the probability might be nonnegligible. Contrast this with a case for which multiple sources of DNA are available. Multiple samples are typically processed independently. If these samples are in agreement, then the probability of a laboratory mix-up is extremely small, as noted by Lempert. If a probability of error is presented, it should be the posterior probability of an error for the case at hand.

Thompson commented on the most notorious of the proficiency tests (the CACLD proficiency test), which I also mentioned in my review article. As Thompson notes, each of three laboratories was asked to com-

pare 50 samples (not pairs of samples), and in each round of proficiency testing one false match was declared by only one company, Cellmark. He goes on to say that the rate of such errors is a function of the number of samples tested rather than the number of pairwise comparisons made; thus he says the error rate is 1/50 whereas I said 1/1225. It turns out that neither of us is correct. Each forensic laboratory was responsible for assessing the relationship between all 50 samples, and two false matches occurred, as can be seen in Thompson and Ford (1991, Figure 8). Therefore, on a very simple level, the error rate for Cellmark was 2/1225. Other laboratories had no false positives, making their estimated rate 0. Nevertheless, these error rates cannot be taken too seriously because of the structure of the CACLD test.

The difficulty in interpreting the CACLD experiment can be illustrated by the following example. Suppose four samples are to be compared and in reality samples *C* and *D* match. If sample *C* were inadvertently loaded twice, once in lane *B* and once in lane *C*, two false positives out of six comparisons would occur. A similar error occurred in the CACLD test. The sample from lane 59 was degraded and a handling error caused the sample from lane 57 to also be loaded onto this lane (the sample from lane 57 was stored in two test tubes). If the sample in lane 59 had not degraded, then extra bands in lane 59 might have alerted the laboratory that an error had occured. The conditions of the CACLD were more difficult than those encountered in routine casework, which detracted from the value of the data. Because of the large number of samples to be compared simultaneously, the samples were handled multiple times, exacerbating the opportunity for error. Routine casework requires that a sample (or a few samples) be compared to a known reference, usually the suspect. Forensic scientists are never expected to make large numbers of comparisons as was required in the CACLD test. A better experimental design would simulate the actual conditions encountered in routine casework.

I agree with Lempert that realistic, blind proficiency testing is unlikely ever to contribute much information if error rates are low, although such testing could have some unforeseen benefits. Nevertheless, no one and no laboratory is capable of being error free. It is important, however, to ask what kinds of errors are likely to occur in the forensic setting. According to the forensic scientists I have spoken to, the error of the suspect being compared to himself, rather than the crime sample, is very unlikely due to laboratory protocol. They argue, based on experience, that a much more likely error is that a suspect is compared to the wrong crime sample. In this situation, the suspect is virtually certain not to match this crime sample.

As Lempert points out, the probability of an incorrect decision based on a laboratory error is greatly diminished by retesting samples at another independent laboratory. However, Lempert does not make it clear that retesting is an option routinely available to the defense. In fact, whenever possible, a portion of the crime scene sample is made available to the defense. Of course, there is always an available sample of any living individuals believed to be involved in the case, such as the suspect or the victim (Wooley and Harmon, 1992). Therefore, the sample can be retested by a second laboratory to determine if a serious laboratory error has occurred. This option is not expensive. It costs only a few hundred dollars, which is borne by the public. The extent to which retesting occurs by the defense is difficult to quantify because, in many states, the results need not be reported. One fact is notable: "defense analyses which have been performed have not been presented in court to impeach the prosecution results, suggest[ing] that the analyses corroborated the first results." (Harmon, 1993)

The defense's ability to retest is often overlooked by commentators because it is presumed (e.g., Lewontin) that the crime scene sample is very limited. This impression may have been created by the much-publicized 1988 *Castro*[4] case. For that case, the evidentiary material was a single fleck of blood on the defendant's watch, and that sample was exhausted by the testing laboratory during its first battery of tests. However, this situation is atypical; generally, there are copious amounts of DNA in a rape or rape–murder case with semen present, and the same is often true for murder cases involving blood from the defendant or victim.

Thompson discusses a handful of cases that he believes exhibits the poor quality of forensic work. It is important to note that the citations are to his own opinions of these cases. These opinions would not necessarily be shared by most molecular experts. He would also have us believe that forensic scientists have free rein when it comes to interpreting the gels, declaring questionable matches and ignoring extra or missing bands as it suits them. The former is presumably cured by the "objective" match criterion, which is, in fact, simply an arbitrary rule that ignores the expected positive correlation between bands. Similar to the likelihood ratio, subjective evaluation frequently leads to a rejection of false positives by incorporating this expectation into one's judgment (Evett, 1993). Moreover, these gels are viewed by many, including the defense experts, the judge and ultimately the jury. When the bands do

[4]*New York v. Castro*, 545 N.Y.S.2d 985 (Supp. 1989).

not meet the so-called objective guidelines required for the declaration of a match or there are extra bands, a justification must be given to the court, which can (and will) be challenged by the defense. Therefore Thompson's argument against summarizing the DNA evidence via a continuous version of the likelihood ratio—"because it would do so by sweeping the underlying issue under the rug rather than addressing it fairly"—is mistaken for two reasons: (i) reporting a likelihood ratio that accounts for the two sources of uncertainty important for most cases, measurement error and bandshifting, would not preclude the defense from challenging the evidence on other grounds; and (ii) a likelihood ratio could be formulated to account for other uncertainty besides measurement error and bandshifting. Consequently, I join with Berry in urging statisticians to develop and evaluate likelihood-ratio-based methods such as those outlined in Section 6 of my review.

Why not combine the probability of obtaining a match with the probability of a laboratory error? The question of whether or not the suspect and crime scene samples are from the same source is distinct and separate from the question of whether or not the correct samples were compared in the laboratory. The evidence is logically presented in two stages: Does the evidence suggest that the samples were obtained from the same individual? If so, is there a harmless reason? A number of explanations are possible, including laboratory error, a frameup, the defendant was at the crime scene, but is not guilty of the crime and so forth. There is no reason why the probability of a match should be bounded by the probability of a laboratory error, as Thompson, Berry and Lempert claim it should be. Proceeding to the logical extreme, suppose the DNA fingerprints were unique. A match could only mean that the same person was measured twice. This result does not preclude the possibility of a harmless explanation for the match, however.

6. SPECIFIC CASES

The points I have been making about testing relatives when relevant and about laboratory error are not based on suppositions. Each point can be illustrated by specific cases. It is my purpose here to describe a few cases in which DNA typing has been used or rejected as evidence, with emphasis placed on how this evidence was interpreted. Whenever possible, I will discuss cases previously raised by the other discussants.

Texas v. Hicks[5] provides an example for which multiple family members were considered suspects. In this case, the grandmother of the defendant was raped and murdered. Hicks was placed by witnesses in the victim's house and in the vicinity of the house at the probable time of the murder. Five other persons had access to the crime scene during this time: a great-great-nephew of the victim, two brothers of the defendant, an unrelated insurance salesman and another person seen in the vicinity. Although Hicks was the prime suspect, the State obtained blood samples from all six individuals mentioned. Only Hicks' DNA matched the crime scene sample. Because the relevant family members were excluded, the prosecution presented a profile probability that did not account for relatives.

Another case involving family members is *Golub v. New York*. In this case, a thirteen-year-old-girl was found raped and murdered in the basement of the home of an unrelated neighbor. Three males resided in the crime scene domicile: the defendant, his thirteen-year-old brother and his father. All three were tested, and only the defendant's DNA matched the crime scene sample. Had the prosecution failed to test these relatives, the defense would have had ample opportunity to undermine the case (and the prosecuting attorney might be looking for a new job). Furthermore, adjusting the likelihood ratio rather than testing family members would have been foolhardy because it would reduce the impact of the evidence against the defendant.

Relatives were implicated in a natural way in the Hicks and Golub cases. Contrast these cases with *California v. Howard*.[6] In this case, there was an abundance of evidence that implicated Howard—notably his wallet, which was found at the crime scene (the victim's home). Howard was a tenant of the victim and had been served eviction papers. Although there may have been reason to suspect others unrelated to the defendant, there was no reason to suspect any relatives of the defendant and therefore none were tested. As a hypothetical point, suppose that Howard's wallet was found at the crime scene but no DNA sample was present. If the victim's blood were found on Howard's personal effects, then the question of relatives would almost surely be irrelevant. For an actual case, see *New Jersey v. Marcus*.

An example of a case in which calculations accounting for relatives and an island population were presented is *Regina v. Ebanks* (Indictment #38/90). The population of this Caribbean island included three brothers and numerous cousins of the defendant. Both the prosecution and the defense presented profile probabilities that accounted for these relatives. Based on this and other evidence, Ebanks was convicted of rape. The case, however, was overturned on

[5] *Texas v. Hicks* (Tex. Ct. Crim. App. No. 70,803, 1993).

[6] *California v. Barney*, 10 Cal. Rptr. 2d 731: (Cal. Ct. App. 1992; Barney and Howard were consolidated in review).

appeal because the court felt that too much weight had been placed on the DNA evidence: the DNA provided strong circumstantial evidence to place Ebanks at the scene, but did not prove that he committed a crime.

A case in which "extra bands" complicated interpretation is *Mississippi v. Parker* (Thompson's Note 1). Although a match was declared in this rape–murder case, faint extra bands were apparent in the crime scene sample. A second individual, the boyfriend of the victim, was tested and his DNA accounted for the other bands on the gel. The defense advanced the theory that the boyfriend committed the rape and murder, not Parker, but this theory was not convincing to the court. The presence of the boyfriend's DNA is another example of evidentiary material placing an individual at the crime scene, but not incriminating him. Because a reasonable explanation was provided by the prosecution for the presence of these extra bands, Thompson's calculations were inappropriate. Sometimes the evidence consists of a mixture of two individuals' DNA, with no explanation for the additional bands. In this instance, a statistical method is available to adjust for the presence of multiple samples, which makes efficient use of the data (Evett, 1987).

Imagine the confusion that would have ensued in *Mississippi v. Parker* if the forensic community followed the advice of Lempert and Thompson. Upon learning that the extra bands matched the boyfriend, the jury would then be given the prior probability of a laboratory error, and the exact probability that a relative of the boyfriend would match this profile. All this despite the verbal communication of the boyfriend that he and the victim had consensual sex. This example illustrates why strict rules for presenting evidence in the courtroom are unworkable.

As discussed by Thompson, an example of the difficulty of an arbitrary matching rule is *New York v. Castro*. Much has been written about the Castro case, but we will focus on only one issue: the difference between the measured sizes of the evidentiary and victim band at the D2S44 locus was 3.06 SD, whereas the Lifecodes arbitrary matching rule is 3 SD. (For an example of a more recent case where the evidence was found inadmissible because 1 out of 10 bands was slightly outside of the 3-SD match window, see *New York v. Keene*[7]). For this and other reasons, there was much debate about whether the DNA matched, and the evidence was found inadmissible. Ironically, after Castro was convicted based on other evidence, he admitted that the blood on his watch was from the victim.

[7]*New York v. Keene*, 156 Misc. 2d 108 (Supp. Ct. Queens Co. 1992).

ADDITIONAL REFERENCES

ALBA, R. D. (1987). Interpreting the parameters of log-linear models. *Sociological Methods and Research* **16** 45–77.

ALBA, R. D. and GOLDEN, R. M. (1986). Patterns of ethnic marriage in the United States. *Social Forces* **65** 202–223.

BALAZS, I. and BUDOWLE, B. (1993). Personal communication.

BALDING, D. J. and DONNELLY, P. (1994). Inference in forensic identification. *J. Roy. Statist. Soc. Ser. A* **157**. To appear.

BEGLEY, S (1993). The meaning of junk. *Newsweek* (March).

BUDOWLE, B., MONSON, K. L., GIUSTI, A. M. (1994). A reassessment of frequency estimates of PVUII generated VNTR profiles in a Finnish, Italian, and a general U.S. Caucasian database: no evidence for ethnic subgroups affecting forensic estimates. Unpublished manuscript.

BUFFERY, C., BURRIDGE, F., GREENHALGH, M., JONES, S. and WILLOT, G. (1991). Allele frequency distributions of four variable number tandem repeat (VNTR) loci in the London area. *Forensic Science International* **52** 53–64.

CAVALLI-SFORZA, L. L. and PIAZZA, A. (1993). Human genomic diversity in Europe: a summary of recent research and prospects for the future. *European Journal of Human Genetics* **1** 3–18.

COLLINS, A. and MORTON, N. E. (1994). Likelihood ratios for DNA identification. *Proc. Nat. Acad. Sci. U.S.A.* To appear.

DAWID, A. P. (1994). The island problem: coherent use of identification evidence. In *Aspects of Uncertainty: A Tribute to D. V. Lindley* (P. R. Freeman and A. F. M. Smith, eds.) 163–174 Wiley, New York.

DEVLIN, B., KRONTIRIS, T. and RISCH, N. (1993). Population genetics of the HRASI minisatellite locus. *American Journal of Human Genetics* **53** 1298–1305.

DONNELLY, P. (1992). Discussion of "Statistical inference in crime investigations using DNA Profiles" by D. A. Berry, I. W. Evett and R. Pinchin. *J. Roy. Statist. Soc. Ser. C* **41** 524–525.

DONNELLY, P. and BALDING, D. J. (1994). The prosecutor's fallacy and DNA evidence. *Criminal Law Review*. To appear.

EVETT, I. W. (1987). On meaningful questions: a two-trace transfer problem. *Journal of the Forensic Science Society* **27** 375–381.

GEISSER, G. and JOHNSON, W. (1993). Testing independence of fragment lengths within VNTR loci. *American Journal of Human Genetics* **53** 1103–1106.

HARMON, R. P. (1993). Legal criticisms of DNA typing: where's the beef? *Journal of Criminal Law and Criminology* **84** 175–185.

HUGHES, J. (1993). Personal communication.

JIANG, C. J. and COCKERHAM, C. C. (1987). Use of the multinomial Dirichlet for analysis of subdivided genetic populations. *Genetics* **115** 363–366.

KAYE, D. H. and KOEHLER, J. J. (1991). Can jurors understand probabilistic evidence? *J. Roy. Statist. Soc. Ser. A* **154** 75–81.

KOEHLER, J. J. (1993). DNA matches and statistics: important questions, surprising answers. *Judicature* **76** 222–229.

KOPEC, A. C. (1970). *The Distribution of Blood Groups in the United Kingdom*. Oxford Univ. Press.

KRANE, D. E., ALLEN, R. W., SAWYER, S. A., PETROV, D. A. and HARTL, D. L. (1992). Genetic differences at four DNA typing loci in Finnish, Italian and mixed Caucasian populations. *Proc. Nat. Acad. Sci. U.S.A.* **89** 10,583–10,587.

LEMPERT, R. (1991). Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. *Cardozo Law Review* **13** 303–341.

LEMPERT, R. (1994). DNA, science and the law: two cheers for the ceiling principle. *Jurimetrics*. To appear.

LIEBERMAN, S. and WATERS, M. (1985). Ethnic mixtures in the United States. *Sociology and Social Research* **70** 43–52.

LOFTUS, E. F. (1975). *Cognitive Psychology* **7** 560–572.

MUELLER, L. D. (1993). The use of DNA typing in forensic science. *Accountability in Research* **3** 1–13.

MURTY, J. S., MURALIDHAR, B., GOUD, J. D., MOHAN RAO, J. RAMESH BABU, B. P. and SRINIVAS RAO, V. (1993). Hierarchical gene diversity and genetic structure of tribal populations of Andhra Pradesh, India. *American Journal of Physical Anthropology* **90** 169–183.

NATIONAL RESEARCH COUNCIL (1992). *DNA Technology in Forensic Science*. National Academy Press, Washington, D.C.

ROEDER, K., ESCOBAR, M., KADANE, J. and BALAZS, I. (1993). Independence in the DNA fingerprint data? A hierarchical approach to the question. Technical report, Carnegie Mellon Univ.

ROTHMAN, E. D., SING, C. F. and TEMPLETON, A. R. (1974). A model for the analysis of population structure. *Genetics* **79** 943–960.

SHIELDS, W. M. (1992). Forensic DNA typing as evidence in criminal proceedings: Some problems and potential solutions. In *Proceedings of the Third International Symposium on Human Identification* 1–50. Promega Corp., Madison, WI.

SLIMOWITZ, J. R. and COHEN, J. E. (1993). Violations of the ceiling principle: exact conditions and statistical evidence. *American Journal of Human Genetics* **53** 314–323.

SPIELMAN, R. S., NEEL, J .V. and LI, F. H. F. (1977). Inbreeding estimation from population data: models, procedures and implications. *Genetics* **85** 355–371.

SUDBURY, A.W., MARINOPOULOS, J. and GUNN, P. (1993). Assessing the evidential value of DNA profiles matching without using the assumption of independent loci. *Journal of the Forensic Science Society* **33** 73–82.

THOMPSON, W. C. (1993). Evaluating the admissibility of new genetic identification tests: Lessons from the "DNA war." *Journal of Criminal Law and Criminology* **84** 701–781.

THOMPSON, W. C. and FORD, S. (1989). DNA typing: acceptance and weight of the new genetic identification tests. *Virginia Law Review* **75** 45–108.

THOMPSON, W. C. and FORD, S. (1991). The meaning of a match: sources of ambiguity in the interpretation of DNA prints. In *Forensic DNA Technology*, (M. Farley and J. Harrington, eds.). Lewis Publishers, Chelsea, MI.

THOMPSON, W. C. and FORD, S. (1993). Is the probative value of forensic DNA evidence undermined by subjectivity in determination of matches? In *Proceedings of the Second International Conference on Forensic Statistics* (W. Harnagel and D. H. Kaye, eds.) A47–A60. Arizona State Univ. Center for Law, Science and Technology.

THOMPSON, W. C. and SCHUMANN, E. L. (1987). Interpretation of statistical evidence in criminal trials. *Law and Human Behavior* **11** 167–187.

TRIBE, L. (1971). Trial by mathematics: precision and ritual in the legal process. *Harvard Law Review* **84** 1329–1393.

WALDMAN, C. (1985). *Atlas of the North American Indian.* Facts on File, New York. (Maps and illustrations by Molly Braun.)

WEIR, B. S. (1993). Tests for independence of VNTR alleles defined as quantile bins. *American Journal of Human Genetics* **53** 1107–1113.

WOOLEY, J. and HARMON, R. P. (1992). The forensic DNA brouhaha: science or debate? *American Journal of Human Genetics* **51** 1164–1165.