Comment: The Use of DNA Profiles in Forensic Contexts

R. C. Lewontin

The article by Roeder raises inter alia the two significant issues in the forensic use of DNA profiles: (1) the reliability of the laboratory results showing a match between the DNA taken from the crime scene and the DNA of the accused and (2) the calculation of the probability that some one other person than the accused person might have been the source of the DNA from the crime scene. Unfortunately, having raised the issues, Roeder fails to come to grips with the real problems that these issues raise. I shall try, briefly, to clarify the situation.

LABORATORY RELIABILITY

Roeder devotes only three short paragraphs to the question of laboratory error and concludes by saying "I believe this emphasis on proficiency testing is misplaced" although no detailed argument is provided. The problem is that Roeder's description of the technique and her reference to paternity testing show that she has missed the essential technical issue. The usual scenario in a criminal trial is that a small amount of dried blood, semen or tissue scraping is recovered from the scene of a crime. A suspect is then identified, and a sample of his or her blood is taken in cubic-centimeter amounts. DNA from the crime scene scrapings and from the suspect's blood sample are then compared side by side in the same laboratory. (Sometimes the tissue or dried blood is found on property of the suspect and it is the victim's DNA that is to be matched with it.) While there is more than enough DNA recoverable from the suspect's large blood sample to carry out the needed procedures, the very small, and often degraded, sample from the crime scene does not contain sufficient DNA for the comparison. To obtain sufficient material, the DNA from the crime scene is "amplified," that is, copied thousands or millions of times in a procedure known as polymerase chain reaction (PCR). As the name of the procedure suggests, the original small number of DNA molecules are copied once, then these copies plus the original are copied a second time and so on for a number of cycles, increasing the total population of molecules exponentially until a sufficient amount has been produced for the matching procedure. The problem with the PCR technique is that because of its chain nature, contaminant molecules in the original sample may also be amplified and, since the original crime scene sample contained so few molecules, contaminants may overwhelm the original in the amplification. In addition, small differences in DNA sequence can have very large effects on the relative amplification of the components in an original mixture.

Now consider the actual practice in a forensic DNA laboratory. A technician is handling two samples. One is the very large DNA sample from the suspect's blood, the other is the minuscule DNA sample from the crime scene, which is then amplified by PCR. The situation is ideal for PCR contamination, with the result that the suspect's DNA will not really be compared with that from the crime scene, but with his or her own DNA that has just been replicated in the PCR reaction. The result will be a perfect match.

All of us who use the PCR technique regularly are acutely conscious of the contamination problem, and the best laboratories have suffered occasionally from it. The perspiration and "oils" on fingertips have provided enough DNA contamination in PCR experiments to give completely artefactual results. Only careful replication catches these errors, and some errors have not been caught until much later when another laboratory found conflicting results. In the forensic context, where the liberty and even life of the suspect is in question, it is essential that courts be assured that laboratories are taking careful precautions against these contamination errors, not to speak of grosser errors of recording etc. Representatives of commercial laboratories that have previously been found to provide erroneous results have told interviewers that they have "cleaned up their act." Perhaps they have, but we cannot know without independent checks, and anyway what about the people convicted before they "cleaned up their act"? The FBI has consistently refused to allow indepen-

---

R. C. Lewontin is Alexander Agassiz Professor of Zoology and Professor of Biology at the Museum of Comparative Zoology at Harvard University, Cambridge, Massachusetts 02138.

dent quality control checks, relying on their own internal procedures.

In a forensic context, where the liberty and even life of a suspect is at stake, there must be frequent, independent and unannounced inspections and tests of DNA laboratories, on the model of the inspections carried out by radiation safety officers and the Department of Energy of laboratories using radioactive materials. The issue is certainly too important to be dismissed by the unsubstantiated opinion of someone unfamiliar with the technical procedures at first hand. If the data themselves are unreliable, questions of probabilities of alternative suspects are irrelevant.

PROBABILITIES OF MATCHES

The second issue, and one to which Roeder devotes considerable attention, is the problem of making probability statements about alternative persons as the sources of the DNA found at the crime scene or on materials associated with the defendant. To make such probability statements it is necessary to have information on the frequencies of different patterns in a population that is to be considered as the source of any alternative to the accused person. It is here that all the difficulties lie, for if different human populations differ nontrivially in their frequency distributions, then assigning a probability of an alternative to the suspect involves deciding how to weight the different populations; that is, we have the problem of assigning prior probabilities. As it turns out, there are two levels of subdivision of the human species at which some differentiation exists. One is the variation among major "races" (Caucasians, Black Africans, Asians, Amerindians, Australian Aborigines, Oceanians). The other is the differentiation between tribal and national groups within those races (between Spaniards and Finns, between Sioux and Yanomami etc.). Unfortunately, Roeder's discussion of the problem is somewhat ambiguous because she uses the term "ethnic group," which can refer to either of these levels. In what follows I will make the distinction explicit.

There is considerable contention and disagreement over whether the differentiation among tribes and nationalities within races is nontrivial from a forensic standpoint. Roeder's view is that there is no important variation at this level, and she martials facts and analyses to support that contention and to refute the claims to the contrary. My own view is the reverse. Each group in this controversy provides facts and analysis that can be taken as evidence for their own viewpoint, and a great deal of literature, cited by Roeder, has been occupied with these struggles. It is not my intention in this article to go over the ground once again, taking each point argued by Roeder and disputing it. The facts necessary to contradict Roeder's position can be found in, for example, Lewontin and Hartl (1991). For our present purpose we can ignore this controversy, because the problem of assigning priors is acute, no matter what the resolution of this particular issue.

There is no contention between the parties on the existence of large frequency differences between major "races." That is why the FBI maintains separate data bases for Caucasians, Blacks, Asians and "Hispanics." The last group is a census category but not a geographical "race" in the sense of the other three since it is an historically recent collection of populations with very different amounts of European, African and Amerindian ancestry. No one contends that these categories should be ignored in constructing data bases, and all parties agree that the differentiation among these groups is as large, if not larger than, the differences among tribes and nations. Indeed, Roeder believes that the between-race component of genetic diversity is twice as great as that between tribes and nationalities. What seems not to be appreciated by Roeder and others of her opinion is that the agreed upon differentiation between the major "races," by itself, is sufficient to bar prosecutors from making statements like "The probability that some other person than the accused could be the source of the blood found at the crime scene is one in 3 million (or one in 80 thousand, etc.)." That is, the problem of assigning priors is no less acute when the differentiation is between major "races" than when it is between tribes and nationalities, although, of course, it becomes even worse when the latter variation exists. To see the force of this problem, let us consider two examples, one hypothetical and one real.

Suppose a dead person, clearly murdered, is found in his or her parked car at the corner of 110th Street and Central Park West in New York. Within several blocks of this intersection there is a large Black neighborhood, a large Hispanic one, and the upper middle class Caucasian population of Central Park West and its side streets. What weighting shall we give these populations? Obviously we would not give each an equal prior probability on population size considerations alone. Suppose we weight them, as suggested by Roeder, in proportion to their numbers. Aside from the problem of deciding how extensively we should draw the boundaries of the populations, there is a hidden prior in such a calculation, namely, that per person each group is equally likely to provide the suspect. However, depending on the circumstances of the case, very different prior weightings would be given to the various groups. If the victim were an upper middle class Caucasian woman, known to be having an illicit love affair, and she had been shot, the prejudices about priors would be quite different than if the victim were a teenage Hispanic, known to be a petty drug pusher, and had died of a knife wound. Yet we have nothing but subjective opinions reinforced by generally shared stories to use in establishing priors, and in no case could we argue that these are objective and "scientific."

Let us turn from the hypothetical case to a real one, cited by Roeder, the fatal assault on a woman in Franklin County, Vermont. The accused, a half Abenaki Indian, whose DNA profile matched the blood from the crime scene, successfully argued that the evidence should not be admitted, because no data base existed on Abenakis and there is well known to be considerable genetic differentiation between Indians and whites, and among Indian tribes. Roeder echoes the argument of Weir and Evett (1992), two strong proponents of the use of DNA typing, who pointed out, quite correctly, that the race of the defendant is not the information needed to establish the priors, but rather the race of potential alternative suspects. Weir and Evett argue that the appropriate data base in the case was "the population," but, unfortunately, they do not tell us what constitutes "the population." Roeder agrees with them that the judge made the wrong decision when he disallowed the evidence. But the judge made the correct decision, because of facts that Roeder does not reveal or discuss, although they are provided in one of the papers cited in her bibliography (Lewontin, 1993).

Franklin County, Vermont, on the boundary with Canada, has the highest concentration of Amerinds, chiefly Abenaki (Abnaki) of any county in the state. This population, partly interbred with French Canadians, lives on both sides of the international border, where they have been concentrated since they were forcibly relocated early in the 19th century. They are a chronically under-employed and unemployed group, at the economic and social margin, with problems of alcoholism and communal violence that accompany such marginalization. Large numbers of this community live in trailer parks, and most of their social contacts are within the group. The victim was an Abenaki, and she was assaulted and killed where she lived, in a trailer park, a large fraction of whose inhabitants were Abenaki. The defense attorney can reasonably argue that it is among the Abenaki that the assailant should be sought, and that an Abenaki data base is, a priori, the obvious one to use for making probability statements. Moreover, the race of the defendant is not totally irrelevant, since the factors that led the police to search for and arrest a person in the Abenaki community are a revelation of the same assumptions about prior probabilities. The prosecution, on the other hand, might reasonably argue that the trailer park is on a major road, that large numbers of people are travelling there and that the general population of Vermont, and perhaps of Quebec, is the appropriate population. Thus, the judgment about priors becomes a question of various opinions on vague social issues, rather than some objectively determined formulation.

Because the assignment of prior probabilities is so nebulous and so subject to arguments about what seems reasonable in the circumstances of the case, the entire project of assigning probabilities of alternative matches is undermined. In an attempt to finesse this problem, the NRC report on forensic uses of DNA recommended the adoption of a so-called ceiling principle. This principle is that the highest frequency of a pattern observed among all the population groups that might reasonably be included should be used in calculating probabilities of alternative matches. This is, in general, biased toward the defendant, although there are frequency distributions for which a multiplication of probabilities of different VNTR's could actually cause the procedure to underestimate the correct probability (Cohen, 1992). Before such a "ceiling principle" can be used, of course, we must have the data bases necessary to choose the highest frequency, so the principle would not have been of any use in the Vermont example.

WHAT IS THE SOLUTION?

First, the recommendations of the NRC report with regard to proficiency testing must be implemented, irrespective of the resolution of other issues. In the absence of independent, random checks, we do not know how often PCR contamination and other errors appear in reported matches, or indeed, in reported mismatches. Yet, the FBI and other enforcement agencies are spending very large amounts of money and time in an attempt to shore up the present system. This includes a large sum to finance a new National Research Council report, in the hope that the second time around the report will be more favorable to DNA profiles. Among the conditions requested by the Director of the FBI is that the new NRC panel not look into reliability testing.

Moreover, in a system of justice in which defendants are to be convicted on the basis of "no reasonable doubt," no guidelines are offered for what error rate would constitute "reasonable doubt." It is the worst possible statistical practice to provide a sophisticated treatment of bad data. Garbage in, garbage out.

Second, the current method of DNA profiling, which requires estimates of probabilities, should be replaced by idiotyping. If, like real fingerprints, profiles were unique to individuals, no probability statements would be required. Such methods are now under development and are in the testing stage. They are based on differences in actual DNA sequence among the "repeats" in the VNTR loci, so that each person can be recognized by a unique signature. If our interest is, indeed, to correctly identify the perpetrators of violent crimes, then it is unclear why we continue to argue about probability calculations and statistical artefacts in place of carrying out the necessary research to create a real "DNA fingerprinting."

Comment

Aidan Sudbury

Imagine a scientific world in which there is no theory of population genetics, but in which a clever technique has been devised which associates with each individual a set of six characteristics. Let us give this technique a name, say, "DNA fingerprinting." It is claimed that these fingerprints identify someone with a high degree of reliability. To test this viewpoint, databases are assembled and it is found that matches are indeed very rare—in fact, that a match between different individuals is found on average 1 in 10,000 times. Compared to other evidence accepted by the courts, such as identity parades, alibis, motives for the crime, this is considered very reliable and has become accepted.

Now, some years later, the theory of population genetics evolves, and a new method of determining match probabilities is based on this theory. Two things may happen. First, the calculations suggest match probabilities of the order of 1 in 10,000. In which case we may say "How interesting! But I don't think we want to burden the courts with the considerable complexities involved with these calculations. We're quite happy with the way we're doing things." Second, the calculations may suggest probabilities of the order of 1 in 100,000, in which case we shall just assume they are wrong.

To return to the real world: it seems that the undoubted charms of population genetics, with its Hardy–Weinberg and linkage equilibrium, have led us into confusing the primary with the secondary evidence. If the observed match probabilities in databases were not small, no amount of testing of databases for independence, or discussion as to just how different allele frequencies are in different races could persuade us that the theory we were using was correct.

As far as I know, all investigations of databases (see, e.g., Risch and Devlin, 1992a, b; Herrin, 1993; Sudbury, Marinopoulos and Gunn, 1993) have shown that matches between unrelated individuals are extremely unlikely. Among related individuals, only the immediate family (brother, sister, father, mother) are sufficiently close to give a probability of a match that is not forensically significant. What perhaps remains to be shown is that matches within small communities are still rare even though there has been a degree of inbreeding in the past. Nichols and Balding (1991) have treated this problem theoretically, but some data covering these situations would be welcome.

Now, let us see how knowledge about the number of matches in a database may be used to make statements about the probability of guilt. Suppose the population can be classified into an unknown number of categories $C_1, \dots, C_n$ and that these have unknown frequencies $p_1, \dots, p_n, \sum p_i = 1$. Further, a sample of size m has been taken and none have been found to be from the same category (there have been no matches). Now, a sample taken from the accused has been found to be in the same category as a crime sample, but both are different from any in the original sample. The aim is to use this data to test the hypothesis H: the accused is innocent.

The probability that the crime samples should match, but no others, under H is

$$(1) \qquad P^* = \sum_{i=1}^{n} p_i^2 \sum_{j_l \neq j_k \neq i} p_{j_1} \cdots p_{j_m}.$$

An appropriate p-value of the test is the maximum of this expression over all sets $\{p_i\}$. Consider the terms involving p_i and p_j. They are of the form

$$(2) \qquad A(p_i^2 p_j + p_j^2 p_i) + B(p_i^2 + p_j^2) + C p_i p_j + D(p_i + p_j),$$

where A, B, C and D are functions of the other p_l. This expression can be written

$$(3) \qquad \begin{gathered}\left[A(p_i + p_j) + C - 2B\right] p_i p_j + B(p_i + p_j)^2 \\ + D(p_i + p_j).\end{gathered}$$

Aidan Sudbury received a Ph.D. in Astrophysics from Monash University, where he is now a Senior Lecturer. His address is: Department of Mathematics, Monash University, Clayton 3168, Australia.