SCIENTIFIC AMERICAN May 1990 Volume 262 Number 5

When Science Takes the Witness Stand

In courts of law, forensic testimony often goes unchallenged by a scientifically naive legal community. Forensic methods must be screened with greater care if justice is to be served

by Peter J. Neufeld and Neville Colman

In the early evening of November 21, 1974, powerful bombs ripped through two pubs in the industrial city of Birmingham, England, leaving 21 dead and 162 injured. The government immediately blamed the Irish Republican Army for the attacks and mounted a massive search for the perpetrators. After a railroad clerk reported that six Irishmen had boarded a train in Birmingham minutes before the first bomb blast, police intercepted the men as they disembarked at the port of Heysham. The six men were taken to the police station, and there, their hands were swabbed with chemicals that would reveal the presence of any nitrites, which would be consistent with the recent handling of explosives. The forensic scientist who performed this procedure, known as the Greiss test, reported positive findings on the right hands of two of the six suspects. That evidence became the linchpin of the government's successful prosecution of the "Birmingham Six."

Now, 16 years later, the six men may be released. The Greiss test, on which their convictions had been largely based, has proved unreliable. It turns out that a variety of common substances such as old playing cards, cigarette packages, lacquer and aerosol spray will, along with explosives, yield a positive result. As it happened, the six men had spent most of their train ride to Heysham playing cards and smoking cigarettes.

The Birmingham case raises troubling issues about the application of forensic technology to criminal investigations. Since the discovery of fingerprinting at the turn of this century, science has assumed an increasingly powerful role in the execution of justice. Indeed, scientific testimony is often the deciding factor for the judicial resolution of civil and criminal cases. The scientific analysis of fingerprints, blood, semen, shreds of clothing, hair, weapons, tire treads and other physical evidence left at the scene of a crime can seem more compelling to a jury than the testimony of eyewitnesses. As one juror put it after a recent trial in Queens, N.Y., "You can't argue with science."

Scientists generally welcome this trend. Because the scientific community polices scientific research, subjecting new theories and findings to peer review and independent verification, it is often assumed the same standards prevail when science is applied to the fact-finding process in a judicial trial. But in reality such controls are absent in a court of law. Instead nonscientists—lawyers, judges and jurors—are called on to evaluate critically the competence of a scientific witness. Frequently lawyers are oblivious of potential flaws in a scientific method or argument and so fail to challenge it. At other times, the adversaries in a case will present opposing expert opinions, leaving it up to a jury of laypersons to decide the merits of the scientific arguments.

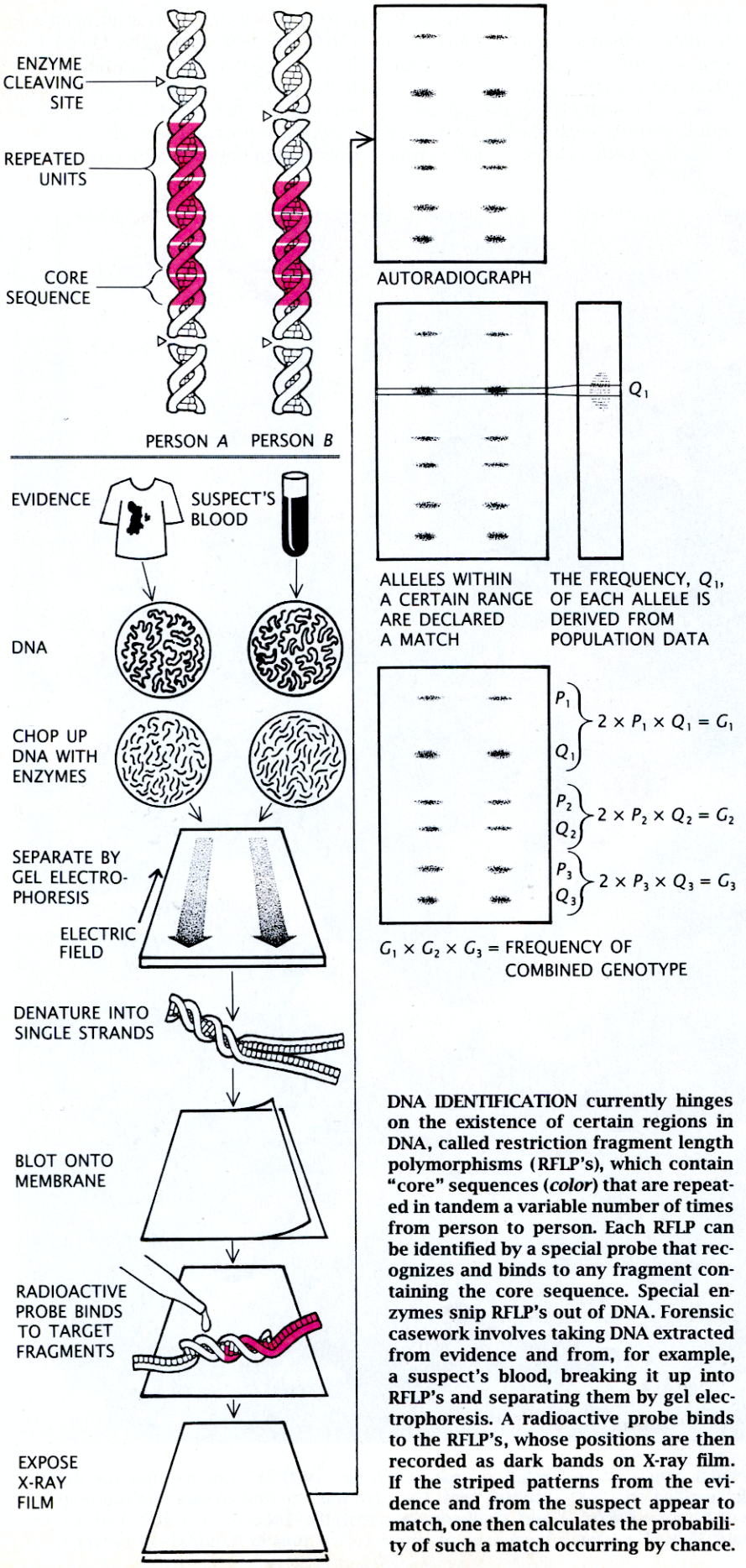
The disjunction between scientific and judicial standards of evidence has allowed novel forensic methods to be used in criminal trials prematurely or without verification. The problem has become painfully apparent in the case of forensic DNA profiling, a recent technique that in theory can identify an individual from his or her DNA with a high degree of certainty. Although

PETER J. NEUFELD and NEVILLE COLMAN have collaborated for several years on the problem of admitting new scientific techniques into criminal cases and have lectured on the subject to both defense attorneys and prosecutors. Neufeld, an attorney specializing in criminal defense and civil-rights litigation, was co-counsel in *People v. Castro,* in which DNA evidence was first successfully challenged. He is a member of the New York State governor's panel on forensic DNA analysis. Neufeld received his J.D. in 1975 from the New York University School of Law and is adjunct associate professor at the Fordham University School of Law. Colman is director of the Center for Clinical Laboratories at Mount Sinai Medical Center in New York City. He received his M.D. in 1969 and his Ph.D. in 1974 from the University of the Witwatersrand, Johannesburg. He has advised counsel and testified in legal proceedings involving the admissibility of scientific evidence.

many aspects of forensic DNA identification have not been adequately examined by the scientific community, police and prosecutors have carried out DNA analysis in more than 1,000 criminal investigations in the U.S. since 1987. Few of these cases reached trial. In most instances, defendants pleaded guilty on advice of counsel after a presumably infallible DNA test declared a match.

Several recent cases have raised serious reservations about the claims made for DNA evidence. Last spring, during a pretrial hearing in *People v. Castro* in New York City, Michael L. Baird of Lifecodes Corporation in Valhalla, N.Y., one of the two major commercial forensic DNA laboratories in the U.S., reported the odds of a random match between a bloodstain and

EXPERT WITNESS Lorraine Flaherty, a molecular geneticist at the New York State Department of Health, testifies on DNA analysis during last year's pretrial hearing of *People v. Castro.* Bronx County Supreme Court Justice Gerald Sheindlin later ruled against admitting key DNA evidence into the double-murder trial. The case was the first to examine thoroughly—and challenge successfully—DNA tests, which had already been used to obtain convictions in hundreds of earlier trials.



DNA IDENTIFICATION currently hinges on the existence of certain regions in DNA, called restriction fragment length polymorphisms (RFLP's), which contain "core" sequences (*color*) that are repeated in tandem a variable number of times from person to person. Each RFLP can be identified by a special probe that recognizes and binds to any fragment containing the core sequence. Special enzymes snip RFLP's out of DNA. Forensic casework involves taking DNA extracted from evidence and from, for example, a suspect's blood, breaking it up into RFLP's and separating them by gel electrophoresis. A radioactive probe binds to the RFLP's, whose positions are then recorded as dark bands on X-ray film. If the striped patterns from the evidence and from the suspect appear to match, one then calculates the probability of such a match occurring by chance.

the suspect at one in 100 million. Eric S. Lander of Harvard University and the Massachusetts Institute of Technology examined the same data and arrived at odds of one in 24. Ultimately, several proponents of DNA testing denounced Lifecodes' data in the case as scientifically unreliable. Some of Lifecodes' key methods were repudiated, casting doubt on the integrity of hundreds of earlier criminal convictions. The ongoing debate over DNA testing underscores the need to deal more effectively with the difficulties that arise whenever complex scientific technology is introduced as evidence in a court of law.

A trial is ideally a search for truth. To help juries in their quest, the law allows qualified experts to testify and express opinions on matters in which they are professionally trained. Yet the esoteric nature of an expert's opinions, together with the jargon and the expert's scholarly credentials, may cast an aura of infallibility over his or her testimony. Hence, to prevent juries from being influenced by questionable evidence or expert testimony, U.S. courts usually review the material in a pretrial hearing or outside the presence of the jury.

To be admitted as evidence, a forensic test should, as a matter of common sense, satisfy three criteria: the underlying scientific theory must be considered valid by the scientific community; the technique itself must be known to be reliable; and the technique must be shown to have been properly applied in the particular case.

The expression of common sense in a court of law, however, is at times elusive. A majority of U.S. courts decide on the admissibility of scientific evidence based on guidelines established in 1923 by *Frye v. U.S.*, in which the Court of Appeals for the District of Columbia affirmed a lower court's decision to exclude evidence derived from a precursor of the polygraph. "Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define," the court declared in *Frye*. "Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."

Judges, scientists, lawyers and legal

scholars have all criticized the *Frye* standard. Some say it is too vague. Some argue that it is unduly restrictive. Still others complain that it is not restrictive enough. Should "general acceptance," for example, require a consensus or a simple majority of scientists? Also, what is it that must be generally accepted? In the case of DNA profiling, is it the theory that no two individuals, except for identical twins, have the same DNA? Is it the various techniques employed in the test, such as Southern blotting and gel electrophoresis? Or is it the specific application of DNA profiling to dried blood and semen samples recovered from the scene of a crime?

Furthermore, what is the appropriate "particular field" in which a technique must be accepted? Does a test for DNA profiling have to be accepted only by forensic serologists, or must it also be recognized by the broader community of human geneticists, hematologists and biochemists? In a recent California case, DNA evidence analyzed by means of the polymerase chain reaction (PCR) was excluded because that method was not generally accepted by forensic scientists. Yet several months earlier a Texas court that was evaluating the identical PCR method looked more broadly to the opinions of molecular biologists and human geneticists and reached the opposite conclusion.

For many applications of science to forensics, the underlying theory is well established, and legal debate rages mainly over whether one must prove only that a technique is generally accepted for scientific research or, more strictly, that the technique is reliable when applied to forensics.

Why the distinction between nonforensic and forensic applications? Scientists commonly accept that when any technology is tried in a different application, such as forensics, it must be tested thoroughly to ensure an empirical understanding of the technique's usefulness and limitations. Indeed, many a technique that has proved reliable for research—polygraphy, for example—has turned out to be of questionable reliability when applied to forensic casework.

Clearly, in order for the courts to evaluate forensic evidence, judges and lawyers must be able to appreciate the scientific issues at hand. Regrettably, lawyers rarely do more than review the qualifications of the expert (typically based on perfunctory queries about institutional affiliation and publications) and verify the facts on which the expert's conclusions are based. The reason for this limited inquiry is simple: most lawyers and judges lack the adequate scientific background to argue or decide the admissibility of expert testimony. Often judges think—mistakenly, in our opinion—that justice is best served by admitting expert testimony into evidence and deferring to the jury for the determination of its weight.

The problem of scientific illiteracy is compounded by the tendency of judges to refuse to reconsider the validity of a particular kind of scientific evidence once it has been accepted by another judge in an earlier case. This practice is founded on the well-recognized need to respect precedent in order to ensure the uniform administration of justice. But in the case of forensic tests, the frequent failure of courts to take a fresh look at the underlying science has been responsible for many a miscarriage of justice.

Perhaps the most notorious example of the problem is the so-called paraffin test (a cousin of the Greiss test employed in the Birmingham Six investigation), which was used by crime laboratories throughout the U.S. to detect nitrite and nitrate residues, presumably from gunpowder, on suspects' hands to show that they had recently fired a gun. The test was first admitted as scientific evidence in a 1936 trial in Pennsylvania. Other states then simply adopted that decision without independently scrutinizing the research.

For the next 25 years innumerable people were convicted with the help of this test. It was not until the mid-1960's that a comprehensive scientific study revealed damning flaws in the paraffin test. In particular, the test gave an unacceptably high number of false positives: substances other than gunpowder that gave a positive reading included urine, tobacco, tobacco ash, fertilizer and colored fingernail polish. In this instance the legal process failed, allowing people accused of crimes to be convicted on evidence that later proved to be worthless.

More recently the debate over scientific courtroom evidence has centered on two applications of biotechnology: protein-marker analysis and DNA identification. Both techniques employ gel electrophoresis to reveal genetic differences, called polymorphisms, in blood proteins and DNA. These two techniques can potentially match blood, semen or other such evidence found at a crime scene to a suspect or victim.

In the late 1960's crime laboratories became interested in protein polymorphisms in populations. The techniques for studying protein polymorphisms were originally developed as tools for population geneticists and were experimentally tested, published in refereed journals and independently verified. The techniques were then modified by and for law-enforcement personnel in order to cope with problems unique to forensic samples, such as their often limited quantity, their unknown age and the presence of unidentified contaminants. These modifications were rarely published in the scientific literature or validated by independent workers.

For example, molecular geneticists study polymorphic proteins in red blood cells and serum by using fresh, liquid blood and analyzing it under controlled laboratory conditions, all subject to scientific peer review. These techniques were then adapted for use on forensic samples of dried blood by the introduction of various modifications, few of which were subjected to comparable scientific scrutiny. No one ever adequately explored the effects of environmental insults to samples, such as heat, humidity, temperature and light. Neither did anyone verify the claim that forensic samples would not be affected significantly by microbes and unknown substances typically found on streets or in carpets.

One of the major modifications made by forensic laboratories was the "multisystem" test. In the original version of this test, three different polymorphic proteins were identified in a single procedure; the purpose was to derive as much information as possible from a small sample. The three-marker multisystem test was further modified by the addition of a fourth protein marker in 1980 by the New York City Medical Examiner's serology laboratory.

By 1987 evidence derived from the "four-in-one" multisystem had been introduced in several hundred criminal prosecutions in New York State. In that year, however, during a pretrial hearing in *People v. Seda,* the director of the New York City laboratory admitted under cross-examination that only one article had been published about that system—and that the article had recommended the test be used only to screen out obvious mismatches because of a flaw that tended to obscure the results.

In *People v. Seda,* the judge ruled that the four-in-one multisystem did not satisfy the *Frye* standard of general acceptance by the scientific commu-

nity and so could not be introduced into evidence. Unfortunately, *Seda* was the first case involving the test in which the defense went to the effort of calling witnesses to challenge the technology. Consequently, the integrity of hundreds of earlier convictions stands in doubt.

In the past two years DNA profiling has all but eclipsed protein markers in forensic identification. The technique is based on a method originally developed to study the inheritance of diseases, both to identify the disease-causing genes in families known to harbor an inherited disease and to predict individual susceptibility when the gene is known.

Crime investigators have embraced the new technique because it offers two significant advantages over conventional protein markers. First, DNA typing can be conducted on much smaller and older samples. And second, DNA typing was reported to offer from three to 10 orders of magnitude greater certainty of a match. Promotional literature distributed by Lifecodes asserts that its test "has the power to identify one individual in the world's population." Not to be outdone, Cellmark Diagnostics in Germantown, Md.—Lifecodes' main competitor—claims that with its method, "the chance that any two people will have the same DNA print is one in 30 billion." Yet, as testimony in the *Castro* case showed, such claims can be dubious.

The hype over DNA typing spreads the impression that a DNA profile identifies the "genetic code" unique to an individual and indeed is as unique as a fingerprint. Actually, because 99 percent of the three billion base pairs in human DNA are identical among all individuals, forensic scientists look for ways to isolate the relatively few variable regions. These regions can be cut out of DNA by restriction enzymes and are called restriction fragment length polymorphisms (RFLP's).

For DNA identification, one wants RFLP's that are highly polymorphic—that is, those that have the greatest number of variants, or alleles, in the population. It turns out that certain regions of human DNA contain "core" sequences that are repeated in tandem, like freight cars of a train. The number of these repeated sequences tends to vary considerably from person to person; one person might have 13 repeated units at that locus, whereas another might have 29. Special restriction enzymes cut DNA into millions of pieces, including fragments that contain the repeated segments. Because the number of repeated segments varies among individuals, so too does the overall length of these fragments vary.

How can these variable fragments be picked out of the haystack of irrelevant DNA segments? The answer lies in "probes" that bind only to fragments containing the core sequence. If the core sequence occurs at only one DNA locus, the probe is called a single-locus probe. If the core sequence occurs at many different loci, the probe is called a multilocus probe. Forensic laboratories currently make use of three different methods of DNA typing: single-locus RFLP, multilocus RFLP and the polymerase chain reaction. Because the single-locus system is the one most widely employed in forensic DNA identification, we will describe it in some detail.

For forensic DNA identification by single-locus RFLP analysis, DNA from various sources is digested with restriction enzymes, placed in separate lanes on an electrophoretic gel and subjected to an electric field. The field pulls fragments down the lane, with smaller fragments traveling faster than larger ones. The fragments, now sorted by size, are denatured into single strands and transferred from the gel onto a nitrocellulose or nylon membrane, which fixes the fragments in place. (Incidentally, anyone who handles nitrocellulose might test positive on the Greiss test!)

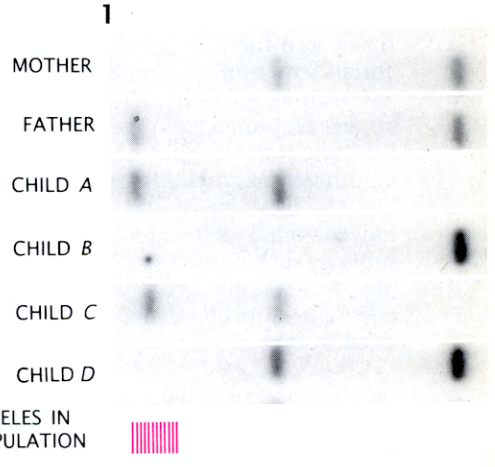
At this point, a radioactive probe is applied, which hybridizes, or binds, to the polymorphic fragments. The mesh is then laid on a sheet of X-ray film to produce an autoradiograph. The radioactively labeled fragments are thereby revealed as a series of bands resembling a railroad track with irregularly spaced ties; the position of the bands is a measure of the size of the polymorphic fragments. The probe can be rinsed away, and a new probe can be applied to identify a different set of alleles.

The autoradiograph resulting from a single-locus probe will ordinarily show alleles of two distinct sizes, one inherited from each parent; such a pattern indicates that the person is heterozygous for that locus. If the probe reveals only one distinct allele, it is assumed that the person inherited the same-size allele from both parents and that the person is homozygous for the locus. Forensic DNA-testing laboratories typically employ several single-locus probes, each of which binds to a different site.

To determine whether two samples of DNA come from a single source, one examines the bands identified by a particular probe on the autoradiograph and decides whether they match. One then refers to data from population-genetics studies to find out how often that particular allele size occurs. A typical allele might be found in 10 percent of the population, making it not all that unlikely that two random people will carry the same allele. But if one looks at alleles at three or four different sites, it becomes increasingly unlikely that two individuals will have the same alleles for all the sites. It is this hypothesis that gives DNA profiling its persuasive power.

How well does forensic DNA profiling stand up under the *Frye* standard? Certainly the underlying theory—that no two people, except for identical twins, have the identical DNA—is unquestioned, and so DNA identification is possible in theory. But is that theory being applied to give a reliable forensic test? And if so, is that test being carried out properly?

In scientific and medical research, DNA typing is most often employed to trace the inheritance of disease-causing alleles within a family. In this diagnostic application, however, one can assume that one allele was inherited from the mother and the other from the father. Because each parent has only two alleles for that gene, barring a mutation, the pattern observed in the child is limited at most to four possible combinations. In addition, if the results are ambiguous, one can rerun the experiment with fresh blood sam-



FORENSIC DNA TYPING is fraught with uncertainty. If the autoradiographs in group 1 are assumed to be from one family, then the alleles of the children must be derived from the parents, even though one of the bands for child *C* is visibly

ples or refer to the alleles of other family members.

In forensic DNA typing, however, it is much more difficult to determine whether an allele from one sample is identical to an allele from another. In the RFLP systems employed in forensics, the number of alleles can run into the hundreds—in contrast to the four from which one must choose when identifying the alleles of a child whose parents are known. Indeed, forensic RFLP systems produce so many different alleles that they virtually form a continuum. In some RFLP's the most common alleles can be crowded into a quarter-inch span on a 13-inch lane. Gel electrophoresis can resolve only a limited number of alleles, however—perhaps between 30 and 100 depending on the particular RFLP—and so alleles that are similar, but not the same, in size may be declared identical. Hence, it can become difficult indeed to declare with confidence that one band matches another. What is worse, forensic samples are often limited in amount and so cannot be retested if ambiguities arise.

These inherent difficulties are further complicated by a problem called band shifting. This phenomenon occurs when DNA fragments migrate at different speeds through separate lanes on a single gel. It has been attributed to a number of factors, involving variables such as the preparation of gels, the concentrations of sample DNA, the amount of salt in the DNA solution and contamination. Band shifting can occur even if the various lanes contain DNA from the same person. Because allele sizes in forensic RFLP systems are closely spaced, it is difficult to know whether the relative positions of bands arise purely from the size of allele fragments or whether band shifting might play a part.

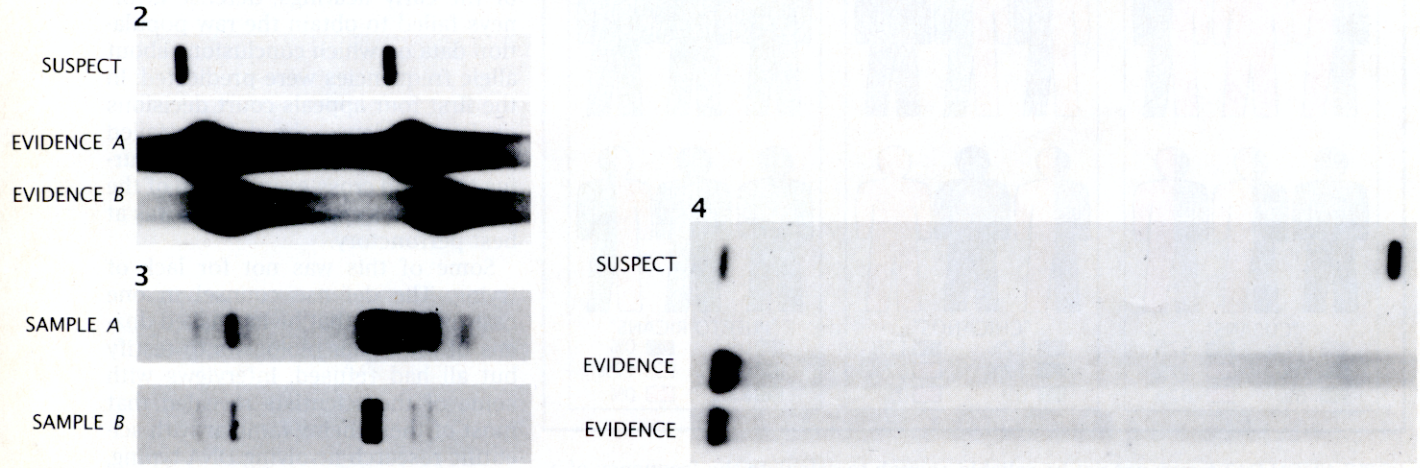
The courts' handling of band shifting is an excellent illustration of the problems that arise when courts, rather than the scientific peer-review process, take on the task of determining whether a method is reliable. Two years ago, when DNA evidence was first introduced in U.S. courtrooms, most forensic DNA scientists rejected the existence of band shifting. But now some experts think band shifting occurs in perhaps 30 percent of forensic DNA tests. There are now many theories about the cause, but as of this writing not one refereed article on the subject has been published.

Forensic DNA laboratories are rushing to develop special probes that bind to monomorphic loci—restriction-enzyme fragments that are the same size in every person—as a possible way to control for band shifting. In theory, if the monomorphic regions are displaced, one would know that band shifting had occurred and could then calculate a correction factor. The difficulty again is that neither this method, nor any other possible solution, has been peer reviewed.

Yet in a rape case tried last December in Maine, *State v. McLeod,* the laboratory director who had supervised the DNA tests for the prosecution testified that a correction factor derived from a monomorphic probe allowed him to declare a match between the suspect's blood and the semen recovered from the victim, even though the bands were visibly shifted. When evidence then came to light that a second monomorphic probe indicated a smaller correction factor, which did not account for the disparity between the bands, he acknowledged that monomorphic probes may yield inconsistent correction factors; nevertheless, he argued that the first correction was appropriate to the bands in question. The prosecutor, though, recognized the folly of defending this argument in the absence of published supporting data and withdrew the DNA evidence. In dozens of other cases, however, judges have been persuaded by the same types of arguments, even though there is no body of research to guide the court. As a matter of common sense, the proper place to first address such issues is in scientific journals, not the courtroom.

Another major problem that arises in forensic DNA typing is contamination. More often than not, crime-scene specimens are contaminated or degraded. The presence of bacteria, organic material or degradation raises the risk of both false positives and false negatives. For example, contamination can degrade DNA so that the larger fragments are destroyed. In such instances a probe that should yield two bands may yield only one (the smaller band).

Research laboratories employ internal controls to avoid the misinterpretation that can result from such artifacts. But such controls may not be suitable for forensic casework. For example, one suggested control for band shifting is to run a mixing experiment: sample *A* is run in lane one, sample *B* in lane two and *A* and *B* in lane three. If



shifted. But if that same lane were of a person whose parentage is unknown, then the band could correspond to one of the other alleles (*color bands*) observed in the population. In group 2, the band patterns from the suspect and from evidence *A* and *B* appear to be displaced relative to one another, which may indicate a band shift. In group 3, sample *A* contains all of the bands from sample *B*, along with extra bands, possibly from contaminants. In group 4, a suspect has two bands, whereas the forensic evidence has only one; the "missing" band may have resulted because degradation of the DNA destroyed the larger fragments. On the other hand, all of these cases could also indicate a real genetic difference.

both samples are from the same person, then ideally lane three would produce one set of bands, whereas if they are from different people, it would show two sets of bands. Unfortunately, in forensic casework there is often not enough material to run a mixing experiment. What is more, recent unpublished studies indicate that certain contaminants, such as dyes, can bind to DNA and alter its mobility in a gel, so that a mixing experiment using samples from the same person can produce two sets of bands.

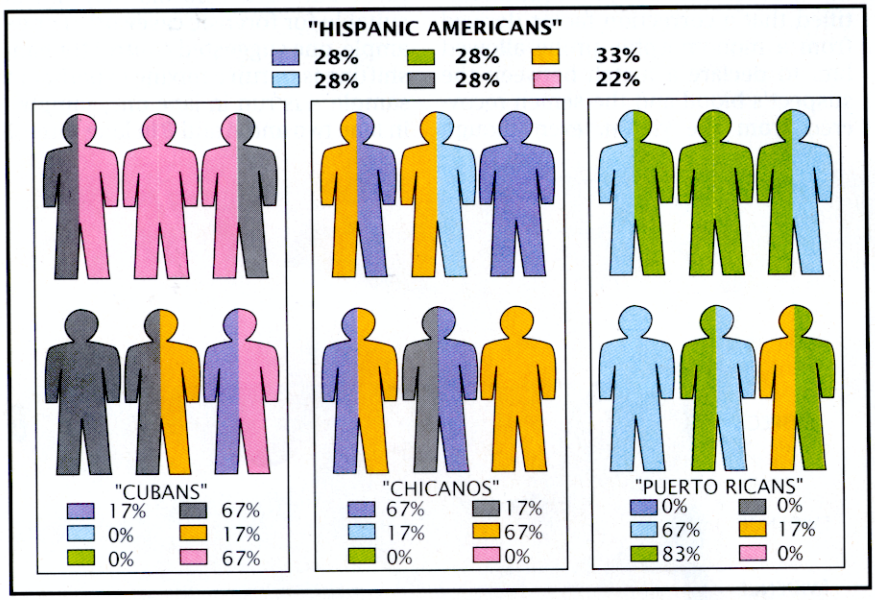
The power of forensic DNA typing arises from its ability not only to demonstrate that two samples exhibit the same pattern but also to suggest that the pattern is extremely rare. The validity of the data and assumptions on which forensic laboratories have been relying to estimate the rarity are currently being debated within the scientific community.

There are two particularly important criticisms. First, because it is difficult to discriminate accurately among the dozens of alleles at a particular locus, the task of calculating the frequency with which each allele appears in the population is inherently compromised. Second, the statistical equations for calculating the frequency of a particular pattern of alleles apply only to a population that has resulted from random mating—a condition that is called Hardy-Weinberg equilibrium.

If a population is in Hardy-Weinberg equilibrium, one can assume allele types are shuffled at random. The occurrence of one allele is then independent of the occurrence of a second allele. One can therefore calculate the frequency of the "genotype," or a particular pair of alleles, for a specific locus by multiplying the frequency of each allele and doubling it (because one has the same probability of inheriting each allele from both parents). The frequency of a genotype for a combination of loci is then obtained simply by multiplying the frequency of the genotype for each individual locus. For example, if the genotypes at loci *A*, *B*, *C* and *D* each occur in 10 percent of the population, then the probability that a person would have these genotypes at all four loci is .1 multiplied by itself four times: .0001.

Forensic DNA laboratories carry out these calculations based on data they have assembled themselves. Most of the data have not been published in peer-review journals or independently validated. One problem is that none of the major laboratories employs the same RFLP system. And even if the laboratories decide to adopt uniform probes and enzymes, the results may still differ significantly unless they all also adopt identical protocols. Commercial DNA-testing laboratories are reluctant to do so, however, because each considers its RFLP system to be proprietary, and the probes and enzymes are sold or licensed to crime laboratories around the country.

Another serious issue is that some populations may not be in equilibrium, in which case neither the alleles nor the various loci may be independent. For such a population, there is as yet no consensus on how to calculate the frequency of a genotype (given the limited data bases of the forensic DNA laboratories). As matters stand, population geneticists are debating whether various racial and ethnic communities exhibit significant population substructures so as to preclude the use of current data bases for the highly polymorphic systems employed in forensic DNA identification. For example, do Hispanics in the U.S. constitute a single mixed population? Or is there nonrandom mating, with Cubans more likely to mate with other Cubans and Chicanos more likely to mate with other Chicanos? Should there be a separate data base on allele frequencies within each of these subpopulations? To find out, population geneticists will need to gather more data.



POPULATION DATA may not yet be reliable enough to calculate the frequency of a genotype accurately. In the hypothetical Hispanic-American population depicted here, a particular DNA site has six distinct alleles, each represented by its own color. Heterozygous individuals are shaded with two colors to represent the two alleles inherited from the parents; homozygous individuals, who have inherited the same allele from both parents, are shaded with one color. Allele frequencies for the entire population differ markedly from allele frequencies for the subgroups shown here.

More than 1,000 criminal investigations in the U.S. have now involved DNA evidence, but in only a few dozen cases has DNA evidence been challenged in a pretrial hearing. According to our own study of these hearings, until the *Castro* case in New York, not one of these hearings addressed the problems of forensic DNA typing that distinguish it from diagnostic DNA typing. In all but two of the early hearings, defense attorneys failed to obtain the raw population data on which conclusions about allele frequencies were predicated. In the first four appeals-court decisions on DNA evidence, the defense failed to present any expert witnesses during trial, and cross-examination of the prosecution's expert witnesses was at best perfunctory.

Some of this was not for lack of trying. The defense counsel in one case explained that he had asked dozens of molecular biologists to testify but all had refused. Interviews with some of the scientists revealed that most of them, being familiar with scientific research involving DNA typing, assumed the forensic application of the technique would be equally reliable. Some who were aware of possible problems were reluctant to criticize the technology publicly for fear that this would be misconstrued as a gen-

eral attack on the underlying science.

Another troubling fact is that defense attorneys are often not able to spend the time or funds required to deal with the complexities of the issues. Novel scientific evidence is most often used to solve violent crimes, and defendants in such cases come predominantly from the less affluent sectors of society. Consequently, most of them must rely on court-appointed counsel selected from public-defender offices, legal-aid societies or the financially less successful members of the private bar. Many of these advocates are exceptionally skillful, but they often lack the time and resources to mount a serious challenge to scientific evidence. And frankly, there are also many less-than-adequate attorneys who are simply overwhelmed by the complexity of the subject.

What is more, in most states a court-appointed lawyer may not retain an expert witness without the approval of the trial judge. In recent DNA cases in Oklahoma and Alabama, for example, the defense did not retain any experts, because the presiding judge had refused to authorize funds. In the *Castro* case, a critical factor in the defense's successful challenge was the participation of several leading scientific experts—most of whom agreed to testify without a fee.

Because defendants are seldom able to challenge novel scientific evidence, we feel that independent overseeing of forensic methods is the only way to ensure justice. Specifically, national standards must be set before a scientific technique can be transferred from the research laboratory to the courtroom, and there must be laws to ensure that these standards are enforced.

The regulation of forensic laboratories has an excellent model: the Clinical Laboratories Improvement Act of 1967 (which was amended in 1988). The act established a system of accreditation and proficiency testing for clinical laboratories that service the medical profession. The law was enacted to ensure that such service laboratories, which are not subject to the same peer scrutiny as research laboratories, would nonetheless provide reliable products and services.

In contrast, no private or public crime laboratory today is regulated by any government agency. Nor is there any mandatory accreditation of forensic laboratories or requirement that they submit to independent proficiency testing. It is also troubling that there are no formally enforced, objective criteria for interpreting forensic data. Four fifths of the forensic laboratories in North America are within police or prosecutor agencies, and so there is an enormous potential for bias because technicians may be aware of the facts of the case. In short, there is more regulation of clinical laboratories that determine whether one has mononucleosis than there is of forensic laboratories able to produce DNA test results that can help send a person to the electric chair.

Accreditation and proficiency testing will work only if implemented with care. National standards for forensic testing must serve the interests of justice, not of parties who have vested interests in the technology. This is not an imaginary danger: from 1988 to 1989 a committee of the American Association of Blood Banks set out to develop national standards for forensic DNA typing and brought in two scientists to provide expertise in molecular genetics; these two happened to be the senior scientists at Lifecodes and Cellmark, the two companies that perform virtually all commercial forensic DNA identification in the U.S.

Some observers suggest delegating the task of setting national standards for forensic DNA identification to the Federal Bureau of Investigation. But there is reason to be wary of this approach. Last year the FBI began to perform forensic DNA identification without first publishing its methodology in refereed journals. In the few pretrial hearings that have challenged DNA tests conducted by the FBI, the bureau has been reluctant to supply the raw data on which it based its criteria, citing its "privilege against self-criticism"—a concept that, incidentally, has little precedent in law. The FBI also opposes independent proficiency testing, arguing that no outsider is qualified to evaluate the bureau's performance. In addition, at a recent FBI-sponsored symposium on DNA typing that attracted 300 forensic scientists from around the country, FBI personnel were alone in opposing proposals requiring laboratories to explain in writing the basis for their conclusions and to have their reports signed by the scientists and technicians who conducted the test.

The FBI's stance on these issues flies against norms established elsewhere in the scientific community. For example, if the author of a scientific article refused to divulge his or her raw data to peer review, the article would be rejected. There is also a clear consensus in favor of independent proficiency tests. If a clinical laboratory refused to comply with any reasonable public request to examine the results of proficiency tests, it would risk losing its accreditation. And it would be unthinkable for a diagnostic laboratory to deliver to the obstetrician of a pregnant woman an unsigned report with only the word "abort" appearing on the page.

Independent scientists are finally beginning to awaken to the urgency of these issues. Last fall the New York State Forensic DNA Analysis Panel proposed detailed requirements for certifying, licensing and accrediting forensic DNA laboratories. The Congressional Office of Technology Assessment is expected to issue a report on the regulation of DNA typing by the time this article appears. The National Academy of Sciences has appointed a committee to study appropriate standards for DNA typing and is expected to issue a report early next year.

It is regrettable that these measures were set in motion only after flaws in current DNA typing came to light in the courtroom. We hope the anticipated reforms will enhance the interests of justice in the future, although this may be small solace to defendants who were wrongfully convicted or to crime victims who saw the true culprit set free. It is our hope that, with appropriate national standards and regulation of forensic laboratories, powerful new forensic techniques such as DNA typing will serve an important and beneficial role in criminal justice. When all is said and done, there should be no better test for identifying a criminal—or for exonerating an innocent suspect.

FURTHER READING

Paul C. Giannelli, THE ADMISSIBILITY OF NOVEL SCIENTIFIC EVIDENCE: *FRYE V. UNITED STATES*, A HALF CENTURY LATER. 80 *Columbia Law Review*, pages 1197-1250; 1980.

Randolph N. Jonakait, WILL BLOOD TELL? GENETIC MARKERS IN CRIMINAL CASES. 31 *Emory Law Journal*, pages 833-912; Fall, 1982.

Black, A UNIFIED THEORY OF SCIENTIFIC EVIDENCE. 56 *Fordham Law Review*, pages 595-695; 1988.

THE CLINICAL LABORATORY IMPROVEMENT AMENDMENTS OF 1988. 42 U.S. Code, Section 263a, 1988.

DNA FINGERPRINTING ON TRIAL. Eric S. Lander in *Nature*, Vol. 339, No. 6225, pages 501-505; June 15, 1989.

DNA TECHNOLOGY AND FORENSIC SCIENCE. Banbury Report 32. George Sensabaugh and Jan Witkowski. Cold Spring Harbor Laboratory, 1989.