University of Michigan Law School
University of Michigan Law School Scholarship Repository

Articles | Faculty Scholarship

2000

DNA as Evidence: Viewing Science through the Prism of the Law

Peter Donnelly
Wellcome Trust Centre for Human Genetics, directorpa@well.ox.ac.uk
Richard D. Friedman
University of Michigan Law School, rdfrdman@umich.edu

Available at: https://repository.law.umich.edu/articles/943

Follow this and additional works at: https://repository.law.umich.edu/articles

 Part of the Courts Commons, Evidence Commons, Science and Technology Law Commons, and the Supreme Court of the United States Commons

Recommended Citation

Friedman, Richard D. "DNA as Evidence: Viewing Science through the Prism of the Law." P. Donnelly, co-author. Law Quad. Notes 43, no. 3 (2000): 87-95. (Adapted from "DNA Database Searches and the Legal Consumption of Scientific Evidence" Mich. L. Rev. 97, no. 4 (1999): 931-84.)

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

The following article is adapted from "DNA Database Searches and the Legal Consumption of Scientific Evidence," 97.4 *Michigan Law Review* 931-984 (1999), and appears here with permission of the Michigan Law Review Association.

DNA AS EVIDENCE:

VIEWING SCIENCE THROUGH THE PRISM OF THE LAW

— By Peter Donnelly and Richard D. Friedman

DNA evidence has transformed the proof of identity in criminal litigation, but it has also introduced daunting problems of statistical analysis into the process. In this article, we analyze a problem related to DNA evidence that is likely to be of great and increasing significance in the near future. This is the problem of whether, and how, to present evidence that the suspect has been identified through a DNA database search. In our view, the two well-known reports on DNA evidence issued by the National Research Council (NRC) have been badly mistaken in their analysis of this problem. The mistakes are significant because the reports have carried great authority with American courts; moreover, the DNA Advisory Board of the FBI has endorsed the second report on this issue. We will also offer some reflections on the habits of mind, of both lawyers and statisticians, that may have led to this result. Part of the problem is a prevailing legal attitude of deference to the scientific establishment. This attitude underlies pending amendments to the Federal Rules of Evidence and pervades the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals Inc.* and its sequels, *General Electric Co. v. Joiner* and *Kumho Tire Co. v. Carmichael*, as well as their precursor, *Frye v. United States*. We will suggest a far less deferential approach.

First, we summarize the nature of DNA evidence and explain the database search problem. We contrast two types of cases. In both, a sample containing DNA has been left, assertedly by the defendant, at the scene of the crime or some other material location. And in both, the defendant's DNA matches that of the crime sample. But the two cases are different in at least one critical respect. In the first case, what we will call the "confirmation case," other evidence has made the defendant a suspect and so warranted testing his DNA. In the other, what we will call the "trawl case," the DNA match itself made the defendant a suspect, and the match was discovered only by searching through a database of previously

obtained DNA samples. The confirmation case has been the model for most uses of DNA evidence up to now. But the recent creation of a national database containing profiles of past offenders means that the trawl case will become increasingly important over the coming years.

Next, we discuss the NRC's analyses of the question of whether and how the DNA evidence should be presented in the trawl case, and we present our own contrasting analysis. The first NRC report (NRC I), in 1992, opposed admission of the evidence altogether. NRC I proposed that a match discovered after a database search should provide the basis for performing further tests on a new sample taken from the suspect identified by the match, but that the evidence of the initial match should not be presented to the jury. The second NRC report (NRC II), issued four years later in 1996, takes a substantially more generous view, but still calls for a testifying expert to drastically understate the value of the evidence.

In our view, the caution reflected in the NRC reports and in the analyses of those who have taken a similar perspective on the problem is unwarranted. The fact that the DNA match was found only after a search, possibly of many thousands of samples, does not diminish the value of the evidence. On the contrary, the fact that other potential suspects have been eliminated as possible sources of the crime sample slightly raises the value of the evidence. No downward adjustment in the force of the evidence is appropriate. (Indeed, the confirmation case and the trawl case are really poles of a continuum, and analysis of the two cases is fundamentally the same.) It may well be that the *total* weight of the evidence is less in the trawl case than in the confirmation case; there may be less inculpatory evidence in the trawl case, and the identification of the defendant in that case may taint any later-developed evidence by suggestiveness. But this is an entirely different matter, and one that can be left to the ordinary process of argument to the jury; it does not require any rules restricting the force ascribed to the DNA evidence or the manner in which that evidence is presented.

Readers not interested in the issues surrounding DNA database searches may skip to the final portion of this article (which begins on page 92). There we argue that the difficulty manifested by the NRC reports has arisen in part from the tendency of statisticians to export to the legal context methods that were developed to assist scientific inquiries and that appear more suitable in that context than in adjudication. But part of the problem also arises from the tendency of courts to defer to the scientific establishment with respect to matters of scientific evidence. We suggest that the solution lies less in a "gatekeeping" role of the type prescribed by *Frye*, as well as by *Daubert*, *Joiner*, and *Kumho*, designed to keep out disreputable expert evidence, and more in the role of aggressive consumer. When courts allow experts to present evidence in court, it is to perform a service for the legal system. The courts should try to ensure that the experts are doing so in a way designed to serve the needs of that system and are not bound by the experts' own professional habits.

DNA evidence and the database search problem

To understand the database search problem, it is necessary to understand some aspects of DNA evidence — but, mercifully, only some of the less technical ones.

DNA is a remarkably complex type of molecule that is sometimes said to contain the genetic blueprint of life. DNA is contained in the nucleus of virtually all cells of every living organism. Within a given organism, the DNA is effectively the same from cell to cell. The entirety of an individual organism's DNA is referred to as its genome. In humans, the genome consists of two collections, one inherited from each parent, of about three billion building blocks, called *bases*. Human DNA is extremely similar across individuals. This shared genetic material is what makes us human and distinguishes us from other life forms. At a multitude of sites in the genome, however, there are variations from one human to another; typically, two unrelated individuals will differ at about one site in 1,000. These variations are what make humans genetically different from each other. Except for identical twins, no two humans have DNA that is identical throughout the whole genome.

The consistency of DNA throughout a given person's body, and the uniqueness of a given person's DNA, are what make DNA evidence so valuable for identification purposes. In determining whether the DNA from two separate samples comes from the same person, it is not possible given the current state of science to compare them over the whole genome. Current testing techniques use several *markers*. Each marker targets a particular place, or *locus*, on the genome. For DNA profiling techniques, loci are chosen that display considerable variability among individuals. In most current methods, this variability is manifested by differences in the length, measured by the number of bases or the number of times a given sequence repeats, between pre-specified locations. This procedure will yield two measurements for each sample for each locus, one for the father's side and one for the mother's side.

If the pair of measurements from one sample at a given locus is the same as the pair of measurements from another sample at that locus, the profiles are said to match at that locus; otherwise, they are said not to match at that locus. If the two profiles match at *each* of the loci examined, the profiles are said to match. If the profiles fail to match at one or more loci, then the profiles do not match, and it is virtually certain — putting aside, as we do throughout this article, the possibility of laboratory error — that the samples do not come from the same person.

Our concern here is with the case in which the profiles do match. A match does *not* mean that the two samples must absolutely have come from the same source; all that can be said is that, so far as the test was able to determine, the two samples were identical, but it is possible for more than one person to have the same profile as indicated by a test even of several loci. At any given locus, the percentage of people having DNA fragments of a given length is

small but not infinitesimal. DNA tests gain their power from the conjunction of matches at each of several loci; it is extremely rare for two samples taken from unrelated individuals to show such congruence over many loci.

But just how rare? That question must be addressed if the strength of the DNA match is to be assessed. Databases of DNA samples from various populations have been collected, and from these it is possible to estimate how common any given fragment length is at a given locus. It is typically assumed that the measurements yielded by each of the markers used in forensic DNA profiling are independent of one another. This assumption enables a forensic scientist to multiply probabilities. This multiplication can yield very low estimates of the probability that a given innocent person from the demographic group described by the database would have a DNA profile matching the profile common to both of these samples. Figures in the range of one in millions down to one in many billions are typical of profiling systems now in use.

So far so good; there are complexities and controversies in the process we have described, but they are not our concern here, and for the most part this process is by now rather well-accepted. Our concern is with the process by which the matching samples came to be tested, and the implications that this has for the value of the evidence and the manner in which it should be presented.

DNA can be used to test for identity in various contexts in litigation, but we will focus on the most important setting. A crime has been committed and a person, assertedly the perpetrator, has left a sample of fluid or tissue containing testable DNA at the scene of the crime or at some other scene associated with the crime. This sample is often known as the "crime sample." The police know that if they find a person whose DNA matches that of the crime sample they may have found the perpetrator. Now we will consider two scenarios by which the police might find a match. In each scenario, we will call the person whose profile matches the crime sample "Matcher."

The Confirmation Case. In the first scenario, what we will call the confirmation case, there is a substantial amount of evidence pointing to Matcher before his DNA is even tested. This evidence might include testimony by a victim of the crime or some other eyewitness identifying Matcher as the perpetrator. It might also include a trail of blood or other circumstantial evidence leading, literally or figuratively, from the crime to Matcher. In any event, the police, believing that they may have their man, secure a DNA test of Matcher. Sure enough, the "suspect sample," as it is often known, matches the crime sample.

Just how this evidence should be presented is a difficult question. One method that is widely used, and that at least for present purposes we regard as satisfactory, states the *match probability*. This is the probability that, if nothing were known about a person other than that he was a member of some defined population, his DNA profile would match that of the crime sample. This probability quantifies the rareness of an "innocent" or "chance" match.

The Trawl Case. In the second scenario — what we are calling the "trawl case" — apart from the crime sample, the police do not at first have evidence that narrows their search to one suspect, or even to a few. This might, for instance, be a rape case in which the rapist was a stranger to the victim and was not apprehended immediately after the crime. But the police do have the crime sample. And they also have a database containing profiles of DNA samples taken from a large number of people, one of whom *might* be the perpetrator. Sure enough, the database search yields one profile, and only one, that matches the crime sample. The police then try to find further evidence incriminating the source of that profile, Matcher, and perhaps they find it. Perhaps, for example, when Matcher is brought before the victim, she identifies him as the perpetrator.

The trawl case will take on increasing importance in coming years. Development has been fast in England and Wales, where since 1995 DNA profiles have been taken routinely from, along others, all persons charged with a "recordable offence." As of July 2000, nearly 80,000 crime samples had been matched to profiles in this database and more than 11,000 matches had been made between samples from different crimes. In the United States, development of a national database has been much more complex. In part because of the encouragement, including financial incentives, offered by the DNA Identification Act of 1994, all 50 states now require designated sets of convicted offenders to provide DNA samples for analysis. The state databases operate in accordance with national quality assurance standards and software designed by the FBI and participate in the Combined DNA Information System (CODIS) maintained by the FBI. Since October 1998, the FBI has been able to compare the profile of a DNA sample from a crime scene with all the profiles in the system. Thus, CODIS now operates in effect as a national database. As of April 2000, laboratories reported analyzing more than 360,000 offender profiles for entry into CODIS, with nearly 400,000 more waiting to be analyzed. As of the same time, CODIS had been responsible for over 600 "hits," assisting in more than 1,100 investigations. These numbers will almost certainly increase dramatically within the next few years.

Over time, one can imagine databases even broader than the current ones; fingerprints are now routinely taken from arrestees as well as convicts, and from many persons not suspected of crime, and it is plausible to suppose that the same will occur with respect to DNA samples. Indeed, the manager of the English database — which operates under fewer administrative and constitutional constraints than its counterpart in the United States — has said he expects that eventually it will include a third of all English men between the ages of 16 and 30.

And now we can perceive what we have called the database search problem. In the confirmation case, the fact that Matcher's sample, the only one tested, matched the crime sample is clearly powerful evidence that Matcher was the source of the crime sample. But in the trawl case, many samples were tested, without a finger already pointing to any particular suspect. How does this

factor affect the strength of the evidence of a DNA match? How, if at all, should that evidence be presented in court? We now turn to that problem.

Analyzing the database search problem

The NRC Reports. In NRC I, a Committee on DNA Technology in Forensic Science specially appointed by the NRC said: "The distinction between finding a match between an evidence sample and a suspect sample and finding a match between an evidence sample and one of many entries in a DNA profile database is important. The chance of finding a match in the second case is considerably higher, because one does not start with a single hypothesis to test (i.e., that the evidence was left by a particular suspect), but instead fishes through the databank, trying out many hypotheses."

Thus, the committee pointed out, "[I]f a pattern has a frequency of 1 in 10,000, there would still be a considerable probability (about 10 percent) of seeing it by chance in a databank of 1,000 people." The report recommended a cautious solution: A match between an "evidence sample" — what we are calling the crime sample — and a profile in a database "should be used only as the basis for further testing using markers at additional loci." That is, evidence of the initial match should be deemed "probable cause" for securing a blood sample from the person so identified, and comparing it with the evidence sample using markers that were not used in the initial test. What is more, if the second test indicates that the two samples match, "only the statistical frequency associated with the additional loci should be presented at trial (to prevent the selection bias that is inherent in searching a databank)."

NRC II offered an analysis of the database search problem that was quite similar to that of NRC I, but recommended a substantially more lenient solution. (Actually, as we explain in our longer *Michigan Law Review* article, NRC II offered two somewhat different analyses; we concentrate here on the analysis to which NRC II gave primacy.) According to NRC II, an "important difference," a difference of a "logical" nature, between the confirmation case and the trawl case is illustrated by this simple set of statistical facts:

> "[I]f we toss 20 reputedly unbiased coins once each, there is roughly one chance in a million that all 20 will show heads. According to standard statistical logic, the occurrence of this highly unlikely event would be regarded as evidence discrediting the hypothesis that the coins are unbiased. But if we repeat this experiment of 20 tosses a large enough number of times, there will be a high probability that all 20 coins will show heads in at least one experiment. In that case, an event of 20 heads would not be unusual and would not in itself be judged as evidence that the coins are biased."

Further, contended the report, "[t]he initial identification of a suspect through a search of a DNA database is analogous to performing the coin-toss experiment many times: A match by chance alone is more likely the larger the number of profiles examined."

Essential to NRC II, therefore, is the perception that the more profiles examined, the less probative the evidence. And NRC II recommends how this supposedly diminished probative value ought to be communicated to the jury. As we have indicated above, an expert may testify as to the probability that a sample taken from an arbitrarily chosen member of the relevant population would match the crime sample. In determining that probability, NRC II provides, the expert should take the probability that she would use if only one sample were compared to the crime sample, as in the confirmation case, and then multiply that probability by the number of profiles in the searched database. If there are hundreds of thousands or even millions of profiles in the database, this adjustment may make what appeared to be an implausible chance of a coincidental match — because the probability was so low — seem to be far more substantial. The DNA Advisory Board of the FBI has endorsed this approach.

Thus, though the NRC reports differ in their ultimate recommendations, their analyses of the database search problem are very similar. We believe that these analyses, and those of scholars who have supported the NRC approach, are clearly wrong. We cannot show here all the anomalies of the NRC approach that we discuss in our longer article. But we will say enough to show that the NRC approach asks the wrong question, and that it fails to recognize the full import of evidence of identification based on a database search.

Our View. The proper view of the situation reflects a rather simple intuition. The value of a DNA match is attributable to the rarity of the profile. If the DNA of a particular person matches the crime sample, that evidence strongly supports the proposition that *that person* was the source of the crime sample; that is, the evidence makes that proposition appear far more probable than it did before the match was known. That other samples have been tested and found not to match does not weaken the probative value of the match, with respect to *this particular proposition*, which is the one of interest at the time of trial. On the contrary, this result somewhat strengthens the probative value of the match, because it eliminates some other persons as potential sources. How probable it appears that the particular person is the source depends not only on the DNA evidence but also on the other evidence in the case. If there is no other evidence pointing to him, then the proposition will not appear as likely as if there were such evidence — not because the DNA evidence is any less valuable, but because the prior probability of the proposition is so low. And evidence found after the DNA match is determined might be subject to a ground of skepticism — the possibility of suggestiveness created by the match itself — not applicable to evidence found beforehand. Thus, the probability that the defendant is the source of the crime sample may well appear less in the trawl case than in the confirmation case, but this is not because the DNA evidence itself is any weaker in the trawl case.

Both NRC I and NRC II emphasized that, as the number of profiles tested increases, so too does the probability of finding a match with the crime sample. That is indisputably true. One can even say that the larger a database is the more likely it is that the database will yield at least one false positive result — a profile that matches the crime scene sample but that does not come from the source of that sample. But the conclusion that the NRC reports draw is that the larger a database is (up to a point) the less valuable is evidence that a database trawl yielded a single match. Here the NRC and its supporters go wrong.

The proposition that the DNA evidence is offered to prove is not the broad one that the source of the crime sample is a person represented in the database. Rather, it is that *one particular person — the defendant in the case at hand — is the source of that sample*. And the evidence bearing on this proposition is not simply that there was one match within the database. Rather, it is that *the DNA of that particular person — alone of all those tested — matches the crime sample.*

Now consider in addition the fact that other samples have in fact been tested and found not to match the crime sample. With respect to the precise proposition at issue — that Matcher is the source of the crime sample — this fact can only *enhance,* not diminish, the probative value of the DNA evidence. One reason for this is that the additional information that a significant number of persons have been tested and found not to match the crime sample can only make the profile of that sample appear rarer than it did absent that information. Potentially more important, a number of people other than the defendant who previously appeared to be possible sources of the crime sample have now been eliminated, thus making each of the remaining possibilities somewhat more probable. Assuming, as is usually the case, that the size of the database is very small in comparison to the suspect population, this effect will be negligible, but as the size of the database increases in comparison to that population, the effect becomes dominant. If the database includes the entire suspect population, then the existence of only one match points the finger without doubt (assuming accurate testing) at the person so identified. This fact alone, that the all-inclusive database makes the existence of one match essentially conclusive evidence, shows that the NRC analysis, which treats the DNA evidence as less valuable the more profiles are in the database, must be mistaken.

The point may be made even clearer by considering an analogy that draws the NRC's hypothetical involving repeated coin flips closer to the reality of DNA testing. Suppose one coin known to be biased, but otherwise indistinguishable from all those in a piggybank, is thrown into the bank. A tester picks from the bank at random a handful of coins and flips each of them 20 times. Each lands heads up approximately 10 times except for one coin that shows heads on all 20 flips. These results are powerful, though not conclusive, evidence that this one coin is the biased coin that was thrown into the bank. Just how powerful the evidence is depends on how common biased coins are believed to be in the bank. But two points seem utterly clear: The evidence that the one 20-heads-up coin is the biased one thrown into the bank is made stronger, not less strong, by the fact that other coins were tried and appeared to be unbiased, and the more other coins that are tested, the stronger the evidence is.

Thus, the DNA evidence itself is *more*, not *less* probative the more profiles have been searched. Nevertheless, given the same DNA match, the *entire body of evidence* may well be stronger in a typical confirmation case than in a trawl case, which involves many searches. There at least two reasons why this is so.

First, by definition, in the confirmation case there is enough evidence independent of the DNA evidence to cast strong suspicion on the eventual defendant. By definition, that is not true before the database search in the trawl case, and it may not be true even after. And it may well be that, even after the trawl identifies the defendant as having DNA matching that of the crime sample, there is little or no other evidence tending to suggest that he is the perpetrator of the crime. Assuming that this is true, and that the database searched in the trawl case was not very large in proportion to the suspect population, the entire body of evidence will plainly be stronger in the confirmation case than in the trawl case. Of course, it may still be strong enough in the trawl case to warrant conviction.

Second, even if the evidence in the two cases, the non-DNA evidence as well as the DNA evidence, is comparable, it may appear to have greater weight in the confirmation case than in the trawl case. Suppose that in the confirmation case the police compile a powerful case against Matcher, based on circumstantial evidence and eyewitness identification, and only at the end of their investigation conduct the DNA test. And suppose that in the trawl case the police, having identified Matcher as a suspect only through a database search, focus their inquiries with such success that they are able to compile the same circumstantial and eyewitness identification evidence. In such a case, it may be plausible that the subsequently discovered evidence was tainted by suggestiveness, given that the DNA match motivated the police, and possibly witnesses as well, to confirm the suggestion that the perpetrator had been found.

These possibilities mean that the confirmation and trawl cases will not necessarily look the same, even assuming that the DNA evidence in the two is of a match with the same profile. But these are factors that a jury, aided by the arguments of counsel, can easily take into account. It is not hard, for example, for a defense attorney to argue, "Except for this DNA evidence, the prosecution does not have a shred of evidence against my client. And the prosecutor's own expert acknowledged that there could well be several other people in the world with the same DNA." Similarly, defense counsel could argue, "The police and the eyewitnesses were hungry to find the perpetrator. After they got this DNA match, naturally they constructed a case to fit the hypothesis of my client's guilt." No technical expertise is necessary to make these arguments, and

no adjustment in presentation of the DNA evidence is required.

The two factors discussed above may account in part for the intuitive sense of some observers — including the NRC committees — that the probative value of a DNA match is weakened by the fact that it is found after a database search. But in fact, as we have argued, it is not that the DNA evidence is weakened; rather, it is possible that the *other* evidence in the case tends to be weaker if, and to some extent because, identification from a database trawl has led to it.

The aspiration to objectivity

We believe there are also deeper reasons for the NRC errors, and for the willingness of legal players to adopt those errors. We believe that some habits of statisticians and scientists make them prone to errors of this sort, and that a judicial tendency towards deference diminishes the ability of courts to make good use of scientific and statistical evidence.

Science aspires to objectivity, to the demonstration of propositions that are not dependent on the subjective views of the observer. Accordingly, it is highly dependent on experiments in which given sets of conditions are observed many times. By counting or measuring different consequences in different conditions, a scientist can hope to draw conclusions on the associations between conditions and consequences.

Accordingly, classical statistics grew up to facilitate objective inferences from data. Classical statisticians try to avoid subjective judgments, seeking instead to determine what conclusions can be drawn solely on the basis of frequency of observation. The Bayesian approach — updating the odds assigned to a given proposition in light of evidence subsequently received — is thus unacceptable to classical statisticians because it depends on the subjective assignment of odds in the absence of objectively measurable data. Instead, the classical statistician, having selected a hypothesis to be tested but without having assigned any probability to that proposition, observes whether the results of an experiment are of a highly unusual nature assuming the truth of that proposition. If they are, then the statistician concludes that the hypothesis can be rejected. Under this approach, great caution must be exercised in testing more than one hypothesis simultaneously, because the more observations are made the more likely it becomes that some will be unusual even though nothing remarkable happened.

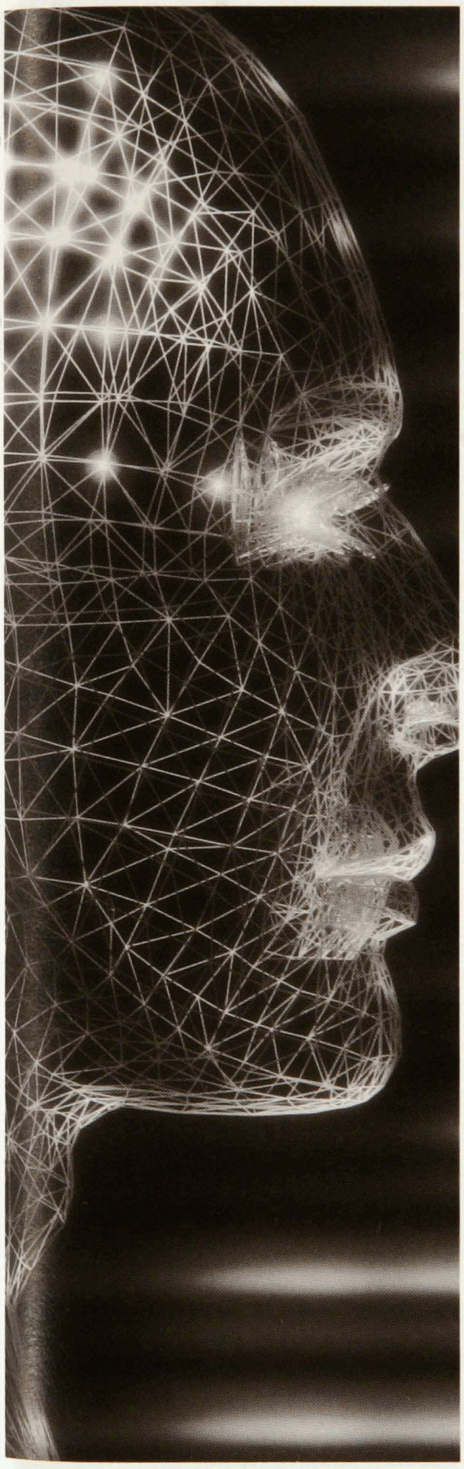
The NRC approach clearly reflects this classical method. NRC I emphasized that with a database trawl "one does not start with a single hypothesis to test (i.e., that the evidence was left by a particular suspect), but instead fishes through the databank, trying out many hypotheses." Thus, no conclusion could be drawn from the fact of a match produced by the trawl; instead, the match "should be used only as the basis for further testing using markers at additional loci." In other words, the initial match yielded by the search identifies the hypothesis to be tested, and nothing more. Though NRC II did not recommend the same solution, its analysis of the problem was very similar. Recall NRC II's discussion of an experiment in which 20 coins are tossed in the air repeatedly and eventually they come up all heads. Applying the lessons of "standard statistical logic," NRC II said that this experiment proves nothing, because given enough trials it is unsurprising that some of them will have unusual results. And this analogy, NRC II maintained, was on point for the database search problem, because the more profiles searched — the more hypotheses tested — the greater the chance of finding one that matches the crime sample purely by coincidence.

The NRC seems not to have recognized how different the enterprise of law is from the scientific enterprise for which the classical statistical model was developed. The problem facing an investigator or a juror is not to determine a general law of the universe. Rather, the investigator must try to determine who committed a given crime on a given occasion; the juror's job is even narrower, seeking to determine whether a given person, on whom its attention has been focused, committed the crime. Subjective assessment, in light of all the information legitimately before them, is essential in allowing both the investigator and the juror to perform their jobs properly. Sufficiently strong evidence may reasonably lead to the conclusion that a given person (whether or not previously identified as a particular source of suspicion) is the perpetrator even though the prior probability of that proposition seemed very low. Once the case is presented to the jury, the jury cannot decide that it lacks the information to draw a conclusion, or that it wants to perform an experiment to ascertain the facts. It must make its best assessment of the facts on the basis of the information within its purview.

Thus, where the evidence tends to show that the person who left a particular DNA sample at a particular place is probably the perpetrator, it becomes of importance to the jury whether the defendant is the source of that sample. The jury should compare how likely the DNA evidence would arise if the defendant were the source to how likely it would arise if he were not the source. And the jury should combine these assessments with its subjective assessments of all the other evidence in the case to assess the probability of guilt.

We have, we think, shown why statisticians are prone in a trawl case to provide the adjudicative system with the answer to the wrong question. But why does the legal system appear to be ready to accept that advice? The answer, we believe, lies in a longstanding tendency of the adjudicative system to defer to the scientific establishment.

Courts have long been afraid that juries will fall prey to "junk science." Thus, they have demanded that an expert witness' opinion be in accordance with a theory that has achieved some threshold level of reputability. For most of the 20th century, the dominant statement of this idea was the



one in *Frye v. United States,* 293 F., D.C. Cir. 1923, that the underlying "scientific principle or discovery . . . from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."

Some states still adhere to the *Frye* test. But many jurisdictions have justifiably come to the conclusion that its demand for "general acceptance" by the relevant scientific community as a pre-condition to admissibility is too stringent. Federal Rule of Evidence 702, which has been adopted in many states, supports this view. The Rule provides:

> "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."

In 1993, in *Daubert v. Merrell Dow Pharmaceuticals Inc.,* 509 U.S. 579, the U.S. Supreme Court held unanimously that the "austere standard" of *Frye* is incompatible not only with the language of Rule 702 but also with the liberal nature of the Rules in general. A majority of the Court went beyond this ruling, however, and attempted to articulate the "gatekeeping role" that the judge must play when "[f]aced with a proffer of expert scientific testimony."

Justice Blackmun's opinion for the majority expressly limited this discussion to evidence based on scientific knowledge, as opposed to "technical, or other specialized knowledge," because that was the nature of the expertise offered in *Daubert* itself. Justice Blackmun put great emphasis on Rule 702's use of the term "scientific . . . knowledge" and operated from the premise that, to qualify as scientific knowledge, "an inference or assertion must be derived by the scientific method." Thus, he attempted to articulate indicia of the scientific method, and he laid out four criteria that should often enter into determining "the scientific validity" of the principles underlying the evidence. These may be referred to as (1) testing, (2) peer review and publication, (3) error rates and standards, and — partially resurrecting *Frye* just a few pages after its apparent death — (4) general acceptance.

Our concern here is not so much with the much-debated questions of whether *Daubert* reflects good philosophy of science, or even whether it reflects good evidentiary policy. Rather, the point of significance here is that, though *Daubert* reflects a loosening of the demands purportedly applied under *Frye,* it still reflects a notable attitude of deference to the scientific establishment. (And this is an attitude that will be entrenched by pending amendments to Rule 702, which barring unforseen intervention by Congress will become effective December 1 of this year and will explicitly require the court to determine whether expert testimony is "the product of reliable principles and methods . . . applied . . . reliably to the facts of the case.") This attitude is apparent in several respects. Most obviously, perhaps, is the continued use of "general acceptance" as a criterion — albeit no longer the exclusive one — for determining admissibility. Further, the other criteria constitute an adoption of currently prevalent scientific methods; the emphasis on peer review and publication also relies heavily on the attitudes of, and decisions made by, the scientific establishment. Perhaps most fundamentally, the entire inquiry seems gratuitous. Under the language of Rule 702, nothing seems to depend on whether the knowledge on which the opinion is based is "scientific" or not. "[S]cientific" and "technical" are clearly listed merely as illustrations of "specialized knowledge"; the key question seems to be whether the opinion is based on "specialized knowledge [that] will assist the trier of fact to understand the evidence or to determine a fact in issue."

The bite of *Daubert*'s insistence on scientific methods was demonstrated in *General Electric Co. v. Joiner,* 522 U.S. 136 (1997). There, the plaintiff attempted to show that exposure to PCBs, a class of chemicals,

could promote cancer. He presented the results of four studies, each of which failed on its own to provide strong support for this conclusion. One of his experts, a toxicologist named David Teitelbaum, testified — somewhat ungrammatically, but comprehensibly enough — at a deposition:

> "[A]s a toxicologist when I look at a study, I am going to require that that study meet the general criteria for methodology and statistical analysis, but that when all of that data is collected and you ask me as a patient, 'Doctor, have I got a risk of getting cancer from this?' That those studies don't answer the question, that I have to put them all together in my mind and look at them in relation to everything I know about the substance and everything I know about the exposure and come to a conclusion. I think when I say, 'To a reasonable medical probability as a medical toxicologist, this substance was a contributing cause . . . to his cancer,' that that is a valid conclusion based on the totality of the evidence presented to me. And I think that that is an appropriate thing for a toxicologist to do, and it has been the basis of diagnosis for several hundred years, anyway."

Justice Stevens agreed with Dr. Teitelbaum that "[i]t is not intrinsically 'unscientific' for experienced professionals to arrive at a conclusion by weighing all available scientific evidence — this is not the sort of 'junk science' with which *Daubert* was concerned." But Justice Stevens stood alone. The rest of the Court upheld the trial court's decision that expert opinions like this one did not rise above "subjective belief or unsupported speculation." A trial court, concluded the majority, could validly decide to exclude "evidence which is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered."

The fear of junk science — the concern that juries will be overwhelmed by chicanery masquerading as science — has thus exerted a powerful effect on the American courts. And perhaps a fear of their own inadequacy to separate the wheat from the chaff has left them to rely greatly on the attitudes and methods of the scientific establishment. We do not mean to deny that a "gatekeeping role" for the courts is necessary; we assume that some evidence offered under the guise of science is of so little value, and sufficient prejudicial potential, that exclusion is warranted. But we believe that deference to the scientific establishment in an attempt to fend off junk science may create another problem — failure to recognize the extent to which, as we have suggested, the methods ordinarily used by scientists do not match up with the needs of the legal system.

The *Joiner* majority's castigation of the toxicologist's opinion as "subjective belief" reflects this failure. Adjudication, we have argued, depends on subjective assessments of factual issues, and courts are fooling themselves if they pretend otherwise. Sometimes, scientists can present to the jury generalized propositions of the type that they try to demonstrate in their ordinary, non-forensic work. But adjudication usually depends on the particulars of the case at hand.

Often this means that the law needs to decide non-recurrent matters for which, because it is impossible to run a controlled experiment or even to gather data across like cases, the scientific method will be useless. And yet, in such cases scientifically based information may be useful in trying to determine the facts. The subjective belief of an expert who has had extensive experience in dealing with problems of a roughly similar nature may be particularly useful in bridging the gap between those principles and the available evidence. This is often the case, for example, in cases in which an engineer offers an explanation for an accident that is similar in some respects to other accidents but unique in some respects. Even if the expert's opinion is not well grounded on scientific principles, her observations and judgment, based on extensive experience, may be useful. For this reason, we find the Court's decision in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), somewhat curious. The expert testimony offered there — that of an expert on tire failure — was clearly not scientific. But the Court held that the trial judge had not abused his discretion in applying the *Daubert* criteria. To the extent that judges apply those criteria in determining the admissibility of evidence that does not even purport to be scientific, *Kumho* will represent a further, and misguided, incursion by science into the realm of law. Fortunately, and appropriately, the Court emphasized that the *Daubert* criteria are not mandatory on the trial court — even with respect to purportedly scientific evidence.

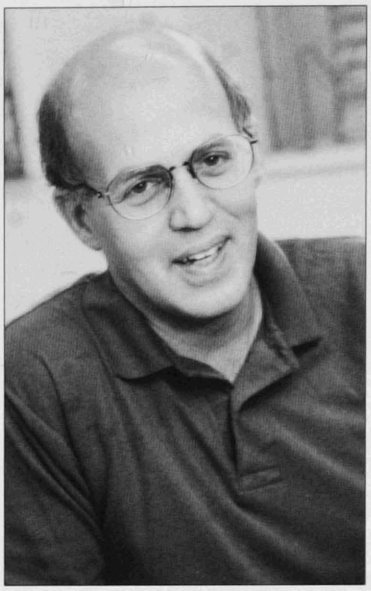
Even if the matter on which an expert wishes to offer an opinion is a recurrent one, so that science can in time yield an answer with confidence, science may not be ready to do so before the legal system needs guidance. Unlike scientific inquiry, Justice Blackmun pointed out in *Daubert*, law "must resolve disputes finally and quickly"; evidentiary rules are "designed not for the exhaustive search for cosmic understanding but for the particularized resolution of legal disputes." But the Court seems to have failed to realize the implications of that perspective. It does not weigh in favor of excluding scientific evidence, or of deferring to the scientific establishment, but rather in favor of recognizing that the law must satisfy its own needs for scientific advice, even if doing so does not square with the usual methods of scientists.

In any case, if a scientist is doing more than reciting general principles without an attempt to relate them to the facts of the case, the law's treatment of scientific evidence must take into account the adjudicative context in which the evidence is offered. That context differs significantly

from the one in which scientists are used to working -– most notably, with respect to evidence of a DNA database search, in that the jury's job is at base a subjective one and in that the bringing of the case tends to define the propositions at issue.

Our perspective may perhaps be crystallized by comparing it to that of Anders Stockmarr, one of the statisticians who has written in support of the NRC approach to the DNA database search problem. "The decision problem of the court," he has written, "should take the implications of statistical hypotheses for data description into account, and not the other way around." In our view, this is precisely wrong. The legal system is a consumer of the information offered by expert witnesses. It may be that the service needed by the legal system requires scientists to operate in ways at variance with their usual operating methods. The law should not be a passive consumer of scientifically based information, taking what scientists have to offer "off the rack." Rather, it should be an aggressive consumer, asking its suppliers to provide what it needs.

This perspective, which has sometimes been apparent in debates concerning psychiatric testimony, may be helpful across the range of expert testimony. What the law needs is not necessarily information processed in the usual ways of science, but rather information that will be helpful to the jurors in making their best subjective assessment of the particular issues at stake in the case at hand. Of course, like any consumer, the law can only ask the supplier to provide what the supplier can. But in the implicit negotiation between law and science, the law has one advantage that most consumers do not have: It sets the rules. The courts should recognize that what they need from science is not the usual output of the scientific community, but rather a special product more tailored to adjudicative needs. Then they may better play the role of aggressive consumer, and so better secure information that will be of help to the trier of fact.



Richard D. Friedman *earned a B.A. and a J.D. from Harvard and a D.Phil. in modern history from Oxford University. His research focuses principally on evidence and U.S. Supreme Court history. He is the general editor of* The New Wigmore, *a multi-volume treatise on evidence, and has been designated to write the volume on the Hughes Court in the* Oliver Wendell Holmes Devise History of the United States Supreme Court. *Professor Friedman clerked for Chief Judge Irving R. Kaufman of the U.S. Court of Appeals for the Second Circuit. He was then an associate for the law firm of Paul, Weiss, Rifkind, Wharton & Garrison in New York City. He came to the Law School as a visiting professor in 1987 from Cardozo Law School and joined the faculty in 1988. Professor Friedman is the Ralph W. Aigler Professor of Law.*

Peter Donnelly *is professor of statistical science and head of the Department of Statistics at the University of Oxford. An Australian by birth, he studied at Oxford as a Rhodes Scholar and earned a D.Phil. in Mathematics. Among his many teaching positions are two stints in the Statistics Department of the University of Michigan, as a visiting assistant professor in 1983-84 and as a visiting professor in 1993. His current research interests relate to applications of statistics in modern genetics, and he has been an expert witness in court cases in both the United States and the United Kingdom.*