Jennifer L. Mnookin

Professor of Law

The connections between science, law and culture inspire Professor Jennifer Mnookin. Her interests include evidence theory, expert evidence and law and culture, with a focus on law and film. Other current work explores the history of expert and visual evidence in the American courtroom. A member of the UCLA School of Law faculty since 2005, she has served as chair of The Association of American Law Schools Section on Evidence and on its executive committee. In 2007, she became the vice dean for faculty and research at UCLA School of Law. Professor Mnookin is a co-author of the treatise *The New Wigmore: Expert Evidence* (with Kaye and Bernstein, 2004). She has published in leading law journals and written op-ed pieces for leading American newspapers.

PEOPLE V. CASTRO: CHALLENGING THE FORENSIC USE OF DNA EVIDENCE†

Jennifer L. Mnookin

In a well-told murder mystery, the reader is left hanging until the very end. Figuring out what actually happened is part of the reader's job, as the author metes out clues and false leads, hints and distractions, bit by bit until it all fits together in a denouement that, ideally, is both surprising and satisfying.

Looking back on *People v. Castro*, there is no particular mystery about what happened. There is no reason to doubt that the defendant, Joseph Castro, a handyman, did in fact commit the murders with which he was charged, the fatal stabbings of Vilma Ponce and her two-year-old daughter. In fact, in *People v. Castro*, there was no trial, for the defendant ended up pleading guilty to second degree murder before the trial began. The case we now call *People v. Castro* was nothing more than a preliminary hearing about the admissibility of evidence at trial. Nor did *People v. Castro* lead to any change in legal rules or to a formal, explicit shift in any evidentiary doctrine.

Why then, should we tell the story of *People v. Castro*? What is it about a preliminary hearing in New York City in 1989 that adduced no new legal standard that has led to the case being cited more than 130 times by later courts and equally often in law review articles? And why is it worth retelling in detail more than 15 years later?

People v. Castro was a preliminary hearing about the admissibility of DNA evidence in the courtroom. It was by no means the first such case—more than a handful of trial courts had already permitted DNA evidence when the preliminary hearing in Castro, that would last 14 weeks and take up more than 5000 transcript pages, began in February, 1989. But *People v. Castro* was, in its own way, an extraordinary drama: the *dramatis personae* were not the defendant or eyewitnesses or the relatives of the victim, but instead, leading research scientists, including winners of MacArthur genius grants and future Nobel Prize winners; forensic biologists who had developed the use of DNA identification for courtroom use and some determined attorneys who, with both grit and luck, managed to put together a set of arguments about the inadequacy of the state's DNA evidence that the judge simply could not ignore. At the preliminary hearing, it was not Joseph Castro who was on trial so much as it was forensic DNA. Much to the surprise of the public, not to mention significant swaths of the legal and scientific community alike, the verdict on new technology that emerged from *Castro* was the Scotch verdict of "not proven."

Although the hearing in *Castro* may not have ultimately made much difference to Joseph Castro, beyond likely reducing the sentence that he was offered in a plea bargain, its broader consequences were significant indeed. *Castro* arguably inaugurated a radical, though perhaps in the end, temporary, shift in the evaluation of DNA evidence. Prior to *Castro*, no court had even come close to rejecting DNA evidence. But *Castro* made it clear that DNA evidence was vulnerable, and enterprising defense attorneys poked and prodded those vulnerabilities in numerous subsequent cases across the country over the next several years, leading a number of courts to reject DNA evidence altogether, something that would have been nearly unthinkable prior to *Castro*.

Castro directly and indirectly affected not only attorneys but scientists, too. *Castro* led to a host of changes of standards in forensic DNA laboratories, and contributed to a set of controversies that motivated additional research, among scientists themselves. Forensic DNA, although no one quite realized it at the time, existed in a tinderbox. *Castro* was the spark that set off a firestorm over the reliability of forensic DNA, a flare-up that grew so heated and intense that press accounts referred to what ensued as "the DNA wars." These battles played out not only in the courtroom but in the pages of leading scientific journals like *Nature* and *Science*. It took a number of years, additional court disputes, continued scientific research and the weighing-in of two distinguished commissions created by the National Research Council, to reach stability and closure in both the legal arena and the scientific one.

People v. Castro is also useful for the insight it provides into the complex intersection of science and law. We can see, on the one hand, how unnatural the adversarial system and its dictates can seem to scientists. We will even see how productive it was, in *Castro*, for the scientists to make an end-run around the adversary system and behave in ways that were both irregular and, in the end, enormously helpful for the production of consensus in the case. *Castro* could therefore be Exhibit A in a sharp critique of the adversarial method of proof, at least with respect to the evaluation of novel scientific techniques. But at the same time, *Castro* also shows how the adversarial system—at least in those instances when parties have equal access to highly qualified experts—may be especially well-suited for revealing limitations and weaknesses in evidence that other evaluative methods, including scientific peer review, reputation and publication, may not necessarily uncover. *Castro* therefore illustrates how the adversarial testing of expert evidence may be *both* truth-obscuring and truth-producing, depending on the circumstances.

Finally, in addition to being an object lesson in the inevitably awkward relationship between legal ways of doing things and scientific ones, *Castro* may perhaps be a beacon. There are significant debates going on right now about many other kinds of forensic evidence, and precisely what and how much evidence of reliability the courts ought to require from them. The substance of the preliminary hearing in *Castro* stands for the idea that the standards of *research scientists* ought to be the standards of *forensic science*—an idea that, if taken to its logical extreme, could make many kinds of commonly-used forensic evidence, from fingerprint identifications to expert document examination to ballistics analysis inadmissible in court until additional research is done to establish the validity of the claims to which forensic experts routinely testify.

In addition to raising these interesting questions about science and law, *People v. Castro* is an interesting story, and it is with the story of the case that we shall begin. After laying out the background facts, I will present an abbreviated history of DNA evidence and its legal use prior to *Castro*, and then describe in some detail the preliminary hearing that made *Castro* special. Then we shall explore the aftermath of *Castro*, concluding with broader ruminations about *Castro*, the use of science in the adversarial system and forensic science.

Facts of the Case

A young man returned to the Bronx apartment he shared with his common-law wife, 20-year-old Vilma Ponce, late in the afternoon on February 5, 1987. He unlocked both of the two locks on the door, but could not enter because the chain locking the door was attached from the inside. He called out the name of his wife and daughter, but was answered only by silence. Concerned and somewhat anxious, he attempted to phone his wife, thinking that perhaps she was sleeping. When no one answered, he called his mother, who lived nearby, to see if she had any possible explanation, but she hadn't spoken to his wife since earlier that afternoon. Growing increasingly concerned, he asked his mother to call the police. He stood outside his building, and attempted to whistle up to his apartment, thinking that maybe his wife or daughter would hear him. Just then he saw a ghastly sight: a man leaving the building, his face, arms and shoes smeared with blood. Moments later, the police arrived. When they entered the apartment, they discovered that Ponce, six months pregnant at the time, and Natasha, the couples' two-year-old daughter, both lay dead, victims of a brutal stabbing. Ponce, found nude from the waist down, had been perforated nearly 60 times, and her small daughter's body had been stabbed at least 16 times. While the victim's boyfriend initially failed to pick Castro out of an array of photographs, he subsequently identified Joseph Castro as the man he saw leaving the building with bloody hands that afternoon. Castro lived nearby, and did odd jobs in various buildings in the neighborhood, including, on occasion, Vilma Ponce's.[1]

Police investigation found further evidence to buttress the eyewitness identification of Castro and to support a circumstantial case that Castro was indeed the murderer. According to one of Vilma Ponce's friends, Ponce had pointed Castro out to her on the street just a week before the murder, complaining that he frequently made suggestive remarks to her. Her friend told her to tell her husband, but Ponce said she didn't want to provoke a possibly violent confrontation between the two men. The police found that one of the locks on Ponce and Rivera's door was improperly installed, and therefore didn't work—and they discovered that Joseph Castro himself, assisting the building superintendent's nephew, had helped to install the malfunctioning lock just two weeks earlier. In addition, because the police found Ponce's just-bought groceries, including meat and chicken, still sitting in a bag on the living room sofa rather than in the refrigerator, they speculated that Ponce had been surprised by her attacker just after getting home—perhaps before she had a chance to latch the second, actually-functioning lock on her door.

All of this was suggestive: it provided the outlines for a story that fingered Castro as a possible suspect and gave tantalizing hints of both motive and opportunity. But the police still might not have had a persuasive case had they not, when they questioned Castro, seized a watch he was wearing, stained with what looked like dried blood. If it *was* blood and if it could be persuasively linked to Vilma or Natasha, that would transform a circumstantial case into a slam-dunk story of Castro's guilt.

The prosecution sent Joseph Castro's watch for DNA testing in the summer of 1987. A few weeks later, Lifecodes, the company that conducted the test, reported that the DNA found on the watch matched Vilma Ponce's DNA profile. Lifecodes claimed that the chance that a person selected at random from the population would match the blood found on that watch was a minuscule 1 in 189,200,000.

The Rise of DNA Evidence

Testing blood for blood types goes back many decades, and over the years ever-more sophisticated tests had been developed. Nonetheless, just a few years earlier, no such definitive identification would have been possible. Although blood testing had grown increasingly sensitive, and had long been able to reliably determine that an individual was *not* the source of a particular blood sample, it could not do any more than show that an individual was a possible source of a blood sample, one among a significant number of people that had the same blood type or the same blood proteins.

Then in 1984, a British scientist named Alec Jeffreys, a DNA researcher at the University of Leicester, made an astounding and surprising discovery. He was studying myoglobin, a protein that stores oxygen, and quite by accident, while working on a problem related to gene mapping, not individual identification, Jeffreys and his colleagues realized that they had found a way to examine a region of DNA that was both inherited in Mendelian fashion (that is, passed along through the generations, half from each parent) and highly variable across individuals. Jeffreys first published the news of his invention in March of 1985, and by July, he and his co-authors were claiming in the prestigious scientific journal *Nature*, that the new technology was a reliable and "unambiguous" way to identify individuals. Quite self-consciously, Jeffreys gave this new technique a name that would resonate: DNA fingerprinting.

Drawing on the widespread belief in the uniqueness and power of fingerprinting was a masterful PR move—it both suggested that DNA evidence shared in the cultural authority of its predecessor, the fingerprint, and it provided for the non-scientific, for those who knew nothing about DNA at all, a mental image of what it was that this new technique could do. As Jeffreys reported in an interview,

> One of the reasons we called this DNA fingerprinting was absolutely deliberate. If we had called this "idiosyncratic Southern blot profiling," nobody would have taken a blind bit of notice. Call it "DNA fingerprinting" and the penny dropped.[2]

Sure enough, the penny *did* drop; the technique quickly got worldwide attention. It was used for the first time in a legal setting that very year, in an immigration dispute in England over whether a teenage boy was in fact a legitimate British citizen returning to be reunited with his British mother, or somebody else, a mere faker who had tampered with a passport. The family's lawyer persuaded Jeffreys to use the new technique to analyze the boy's DNA, and the tests results identified the boy as the mother's biological son. In the face of this DNA evidence and under pressure from the appellate tribunal, the British Home office ended up withdrawing their case. This saved the tribunal from having to decide about the admissibility and validity of a powerful but untested brand new technology. Press accounts nonetheless celebrated both the result and the new technique.[3]

Shortly thereafter, in 1986, the technique was put to use for the first time in a criminal investigation.[4] A teenage girl had been raped and murdered in rural England in 1983, and then in 1986, another girl was found dead nearby. A kitchen worker with a low IQ was fingered as a possible suspect. The DNA evidence from both murders was tested against the suspects, and it turned out that the two criminal samples matched each other, strongly suggesting that the same person committed both crimes. But much to the disappointment of police investigators, neither sample matched the original suspect. The kitchen porter was freed, the first criminal suspect in history to be exonerated by his own DNA. Frustrated and desperate for new leads, police eventually decided to take on a genetic manhunt: every man in the appropriate age range in the vicinity was asked voluntarily to submit blood for testing. Although thousands of samples were tested, none matched the DNA evidence extracted from the semen found at the murder sites. An enormous and controversial effort seemed to have produced nothing useful. Then the police got a much-needed lead: it turned out that a young man who worked at a bakery let slip that he had been coaxed into giving blood in place of one of his co-workers. Police unraveled the story and confronted the co-worker, who promptly confessed to the murders. The DNA evidence confirmed the confession: this time, the police had found the killer.

By 1987, Jeffreys and his research institute had sold their rights in their DNA technique to ICI, a private company that would further develop and commercialize their new technique. That same year, ICI opened up Cellmark Diagnostics USA, and began to offer the technique for paternity testing and forensic matters in the United States. Around the same time, another company, Lifecodes, began offering its services for forensic analysis—based on somewhat different techniques for analyzing DNA, but also designed to provide reliable information about whether two biological samples were likely to have come from the same source.[5]

The DNA techniques used by both companies measured the length of particular genes. DNA profiling looks at specific parts of the human DNA with no known function (therefore sometimes called "junk DNA"). At these places, or *loci*, on the genome, short sequences of DNA are repeated, but the number of repetitions is highly variable across the population. (These are known as VNTRs, or "variable number tandem repeats.") Each possible variant is called an *allele*. Every person inherits DNA from both parents, so an individual will typically have two alleles at each locus. At any one locus, two people could easily have the same allele, but if you examined their alleles at several different loci, the chance that all their alleles would match decreases exponentially.

To examine DNA with these methods, it is first chopped into small pieces using special enzymes that break it apart whenever certain patterns of base pairs within the DNA are found. This broken-up DNA is next divided up using a technique called "electrophoresis," in which the DNA is loaded into a lane in a gelatin slab. In the forensic context, the DNA found at the crime scene would be loaded into one lane, while the DNA known to come from, say, the victim and the suspect are each loaded into their own lanes. All the lanes are then subjected to an electrical current, and because they are different sizes, they travel at different speeds down the gel. Then the DNA is converted from double-stranded to single stranded, transferred and affixed to a membrane and exposed to a radioactive "probe" that latches onto it and can be visualized by exposing x-ray film to the membrane. This produced an autoradiogram, or "autorad," which visually displays the bands attributable to each allele at each locus. If two DNA samples came from the same person, the bands displayed on the autorad would line up and show a "match."

Both companies advertised their new techniques in very strong terms. One advertisement in *Trial* magazine proclaimed "Only DNA Fingerprinting Determines Paternity in Just Two Words: Yes / No. Thirty billion to one accuracy in one conclusive test." "NO ifs, NO maybes," announced another advertisement. In *Criminal Justice* magazine, around the time of the *Castro* case itself, one of Cellmark's advertisements showed two cuffed hands linked by a chain in the shape of the DNA double helix, explaining that DNA fingerprinting "positively identifies suspects . . . by examining a suspect's one-of-a-kind genetic material."

The Early Legal Reception of DNA Evidence

These advertisements—making strong claims for the technology and broadcasting total confidence in its results—illuminate not only how DNA analysis was portrayed by its developers, but also the context in which the first judges in the American courtroom received the exciting new technique. It was presented—not only in these advertisements, but to some extent in the early court cases as well, as a kind of "magic bullet"—powerful, infallible, almost miraculous.

In a number of the earliest court cases involving DNA, scrambling defense lawyers were unable to find any expert witnesses of their own. For example, in *Andrews v. State*, the prosecution hired as a consultant and expert witness David Housman, a prominent biologist from MIT. The defense, by contrast, had no expert witnesses.[6] Later, the defense attorney in the case explained that although he had made calls to many biology departments, he was unable to find anyone interested in getting involved—and many scientists had told him that if Housman was standing behind the evidence, than it was almost certainly valid.

According to historian of science Jay Aronson, this problem grew even more acute over the next year.[7] In addition to the impressively-credentialed molecular biologists who worked for the DNA testing companies, prosecutors soon had at their disposal a growing list of highly respected academic researchers who were prepared to testify in favor of the new technology. Very prominent scientists—geneticist Kenneth Kidd from Yale, molecular biologist Richard Roberts (who would go on to win the Nobel prize in biology in 1993), and several others of equally high repute—testified about the principles of molecular biology and population genetics, and affirmed the legitimacy and validity of the DNA identification techniques. The early defense witnesses—when there were any at all—were not nearly as prominent. Interestingly, the prominent scientists testifying for the prosecution, while leading experts in DNA techniques in general, had very little knowledge about the *forensic* use of DNA or what distinctive problems might arise in the forensic identification context. Questions of *technology transfer*—the special problems that might arise in translating DNA testing from the research laboratory or clinical setting into the forensic science context—did not strike these experts as either salient or problematic. Nor had they examined in detail the specific probes and validation techniques used by the DNA profiling companies, nor the data underlying their population genetics. In fact, both the probes that the companies were using and their population databases were deemed by the companies to be proprietary knowledge, trade secrets.

Given this state of affairs, it is not all that surprising that the early judicial opinions about DNA evidence not only deemed the technique admissible, but sometimes engaged in rhetoric that borders on the reverential. For example, in *People v. Wesley*, the first trial judge in New York State to consider the admissibility of DNA evidence wrote:

> The immediate advantage of DNA fingerprinting … is the claimed certainty of identification. Blood-grouping identification tests often can narrow down the number of suspects to from 30 to 40% of the population. The laboratory the People propose to utilize claims a mean power of certainty of identification for American Whites of 1 in 840,000,000; for American Blacks, 1 in 1.4 billion. There are approximately only five billion people in the entire world.
>
> The overwhelming enormity of these figures, if DNA fingerprinting proves acceptable in criminal courts, will revolutionize the administration of criminal justice. Where applicable, it would reduce to insignificance the standard alibi defense. In the area of eyewitness testimony, which has been claimed to be responsible for more miscarriages of justice than any other type of evidence,

> again, where applicable, DNA fingerprinting would tend to reduce the importance of eyewitness testimony. And in the area of clogged calendars and the conservation of judicial resources, DNA fingerprinting, if accepted, will revolutionize the disposition of criminal cases. In short, if DNA fingerprinting works and receives evidentiary acceptance, it can constitute the single greatest advance in the "search for truth", and the goal of convicting the guilty and acquitting the innocent, since the advent of cross-examination.[8]

New York, like most states at that time, evaluated novel forms of scientific evidence under the *Frye* standard, named for a 1923 case in which a systolic blood pressure test, an early and crude attempt at a lie detector, was excluded from evidence.[9] Admissibility under this test depended on whether the new form of science was "generally accepted" by the relevant scientific communities. Sure enough, after a detailed review of both the substance of the testimony and the glowing credentials of the prosecution experts, the trial judge in *Wesley* found that DNA fingerprinting was "reliable and has gained general acceptance in the scientific community," and hence, was admissible in court.[10]

The judge in *Wesley* was especially reassured by his belief that DNA evidence could never provide an erroneous result. He wrote:

> A matter of extreme significance testified to by Dr. Roberts, and confirmed by [the other prosecution experts] and unrefuted by the defense experts, is that it is impossible under the scientific principles, technology and procedures of DNA fingerprinting (outside of an identical twin), to get a "false positive"— i.e., to identify the wrong individual as the contributor of the DNA being tested. If there were insufficient DNA for the test, or if the test, or any of its steps, were performed improperly, no result at all would be registered—in other words, the autoradiograph would be blank. Thus the dichotomy can never be between an accurate answer and a false answer, but only between an accurate answer and "no answer." Under the undisputed testimony received at the hearing, no "wrong" person, within the established powers of identity for the test, can be identified by the DNA fingerprinting test.

This assumption—that interpreting a DNA test was a straightforward process that would inevitably provide either the right answer or no answer at all—was shared by many of the early courts that considered the admissibility of DNA evidence. As we shall see, after *Castro*, this belief was no longer tenable.

Meanwhile, at the end of November, 1988, an academic conference was taking place that would turn out to have significant consequences for the *Castro* case. Some molecular biologists and forensic scientists had decided to organize a meeting at the renowned Banbury Center to create the opportunity for a wide variety of participants in DNA typing—including molecular biologists, forensic scientists, population geneticists, lawyers and judges, to discuss DNA techniques, their power and their limits. This meeting was the first structured occasion for discussion between university-based academic molecular biologists and commercially employed forensic scientists, and by all accounts, the conversations were both lively and contentious.

One of the participants at the Banbury Conference was Eric Lander, a brilliant MIT scientist who had received a MacArthur "genius" grant one year earlier for his work on techniques to help decipher the human genome. At the Banbury conference, he presented a paper suggesting that the population genetics and statistical issues surrounding the interpretation of DNA profiling—essentially, the knowledge necessary for making the claim of a 1-in-something chance that a random person would have matched the biological material in question—were significantly more complicated than the private companies had acknowledged. Over the course of the meeting, his probing remarks combined with his critical perspective caught the attention of two defense attorneys participating in the conference, Barry Scheck and Peter Neufeld. Both were members of a New York State panel commissioned to study the forensic use of DNA, and they had also recently taken over Joseph Castro's defense from a court-appointed lawyer who felt overwhelmed by the DNA evidence. Scheck and Neufeld are now very well known both for their creation of the Innocence Project, which uses post-conviction DNA evidence to exonerate the wrongly convicted, and for their role as members of O.J. Simpson's defense team. At the time of the Banbury conference, Scheck was a clinical law professor at Cardozo and Neufeld a sole practitioner, and they were beginning to look for experts who might be able to help them challenge the DNA evidence in the case.

Toward the end of the conference, Peter Neufeld approached Eric Lander and asked him to take a look at the DNA evidence in the *Castro* case. As Neufeld told the story to the press after the *Castro* case was over, Lander took a look at the autorad and said,

> "Let me show you how we do things in science." Lander then called over several colleagues, slapped the autorad up against a window, and said, "Match, or no match?"
> "Garbage," one responded.
> "Do it again," said another.
> "Garbage," said a third.[11]

Neufeld asked Lander if he would testify for the defense in the *Castro* case. He declined. He had plenty on his plate already, and besides, at some level, he doubted that there were serious problems with Lifecodes' methods or their implementation. But he did agree to help educate Scheck and Neufeld about DNA evidence and to assist them in making effective discovery requests in the case.

The more Lander learned about Lifecodes' practices in this case, the more disturbed he became. By the time the *Castro* case was over, Landers had agreed to testify after all, had written a 50-page report for the defense and had devoted more than 350 hours of his time, all of it pro bono, to the case.

People v. Castro: the Preliminary Hearing[12]

By the time the *Castro* hearing began in February 1989, DNA had been used in quite a few cases throughout the country; some knowledgeable participants estimated that it might already have been used in as many as 80 proceedings nationwide. In many more cases defendants had accepted plea bargains in the face of DNA evidence. Up until *Castro*, every judge confronted with DNA evidence had deemed it admissible. However, these cases had resulted in only a handful of written opinions and even fewer appellate opinions.

The *Castro* hearing—which lasted more than three months—was presided over by Judge Gerald Sheindlin, who would thereafter retain an abiding interest in DNA. He later wrote two books relating to DNA: *Blood Trail: True Crime Mysteries Solved by DNA Detectives*, and *Genetic Fingerprinting*: *The Law and Science of DNA Evidence*, both published in 1996, as well as a never-produced screenplay about a murder case in which DNA evidence played a role. Sheindlin also did a stint from 1999-2001 as a television jurist on The People's Court. (Sheindlin's wife Judy, at the time of the *Castro* hearing, was a Supervising Judge on the Manhattan Family Court—but she is much better known today as television's Judge Judy.)

The prosecution, led by Risa Sugarman, the homicide bureau chief for the Bronx district attorney's office, began its case by offering the testimony of geneticist Richard Roberts, who explained DNA typing and told the court that it was indeed generally accepted in the scientific community. The next witness was Michael Baird, Lifecodes corporation's chief scientist, who described the techniques and methods Lifecodes used to get results. Up to this point, the preliminary hearing seemed to be business as usual, not particularly different from, say, the prosecution's evidence presented in the *Wesley* case a year earlier.

But then Eric Lander entered the picture. Michael Baird later told an interviewer,

> "Things were going pretty routinely in terms of presenting the background, presenting the data, presenting the information. Suddenly Eric Lander shows up for the defense and has a booklet that is numerous pages thick that has what he critiques as all kinds of problems with the case. The prosecutor in that case is like, 'Who is this guy? Where did he come from?' . . . You know Scheck and Neufeld spent half a day just on his credentials to show that this guy walks on water before the judge."[13]

Indeed, Lander's participation in *Castro* marked the first time that the defense counsel had an expert witness every bit as illustrious as those offered by the prosecution.

But well beyond Lander's sterling credentials, Scheck and Neufeld were able to present a number of extremely significant challenges to Lifecodes' DNA evidence. In brief, the defense arguments fit into three categories: (1) that Lifecodes had failed to follow their own protocols for both declaring a match and interpreting its probability; (2) that forensic analysis posed challenges different from DNA analysis in the research setting, and that these technically demanding challenges posed by "technology transfer" had not yet been fully met by Lifecodes, as indicated by some of their analysis and interpretation of the blood evidence in this case and (3) that there were additional problems with the population genetics databases that Lifecodes was using to determine the probability of a "match."

Sloppiness in Declaring a Match

In Lifecodes' report of their test of the DNA found on Castro's watch, they stated that they were able to test the blood at three loci. At each of these loci, they reported a precise match between Vilma Ponce's blood and the stain found on Castro's watch. At one of the loci, the D2S44 locus, they reported that both samples were homozygous (in other words, had two identical alleles, or just one band on the autorad) and had a band sized at 10.25 kb (Kb stands for kilobase, a unit of size measuring 1000 base pairs on the DNA ladder). But when Lander and the attorneys looked over the materials they had received through discovery, they learned that, in fact, the band in Ponce's blood was actually 10.35 kb, and the band from the bloody watch measured 10.16 kb. (The reported 10.25 was, in fact, the average of the two measurements.) Small variations in measurement were nothing uncommon—the hard question was how much discrepancy in measured size could still support the claim that the two bands really did "match."

In published papers, Lifecodes had maintained that technicians first made matches "visually"—in other words, they "eyeballed" them to see if they looked the same. But they also claimed that visual matches were confirmed through computerized comparison, and their own protocols required that the bands size be within three standard deviations of each other in order to call it a match. But Lander found that 10.16 and 10.35 differed from each other by more than three standard deviations! And this wasn't the only example of interpretive overreaching. Similarly, he found that one of the bands on another of Ponce's loci differed from the "matching" band on the bloodstain by more than 3 standard deviations. In other words, if Lifecodes had followed their own published standards about when to declare a match, they actually should have said that the blood from the watch and Vilma Ponce's blood did *not* definitively match.

On cross-examination, Michael Baird was forced to acknowledge that, no matter what the published papers said, in fact, Lifecodes technicians often just identified a match solely through visual observation, rather than using any predetermined objective standards. Scheck was pleased by this damaging concession, and lifted his arms "in a touchdown-like motion," to the irritation of the district attorney. On redirect, Baird pointed out that all of the measured bands were within three standard deviations of the *average* measurement. When Lander testified later in the hearing, he mocked Baird's effort to recover: "That's similar to saying New York and Boston are both within one hundred and twenty-five miles [of] each other because they're each within one hundred and twenty-five miles of Hartford. I found it somewhat difficult to take seriously."

Even more disturbing, while Lifecodes was clearly a little loose about the measurements required for declaring a match, when it came time to determine the probability of a match, they used a different matching rule, one much stricter than the 3 standard deviations. In essence, they appeared to be using one set of rules for deciding whether a match existed, and another, stricter rule for determining the statistical likelihood that two samples matched. To illustrate the problem with this disparity, suppose that I wanted to know how many law students at a particular school were both 24 years old and had summer birthdays. In order to decide who counted, suppose I defined "summer" broadly and counted all 24 year olds with birthdays anytime between March and October, inclusive. But then in order to decide how likely it was that someone I picked at random from the law school community would "match" my criteria—that is, be 24 and have a summer birthday—suppose I now only counted those people with birthdays between June and August. Using these narrower criteria, I might find that only 1 in 25 people picked at random matched. But using the broader criteria, I might find instead that 1 in 10 matched. The misleading aspect is to use *one* criterion for determining what counts as a match and a *different* one for interpreting the probability of a match—and yet that is exactly what Lifecodes seemed to have done. As Lander put it in his testimony:

> Whatever choice you make for your matching rule, when you go and tell a court what is the chance this would have arisen at random in the population, you had better be using the same matching rule. To do otherwise is to report a probability that is simply not true. If I go out and I catch matches with a ten-foot wide butterfly net and I say I caught a match, and then I come to court and I say, and it was so rare that I caught this match, and I will prove it to you by showing that when I go out with a six inch butterfly net, I never catch matches in the population, that would be absurd.

The defense, by looking closely at the actual data on which Lifecodes' claims were made, was therefore able to show significant problems with Lifecodes' implementations of their own procedures.

The Special Problems of Forensic Science

In the research and diagnostic setting, and with paternity testing as well, the issues involved with using DNA are more straightforward. Blood is plentiful. It can be taken under sterile conditions, and kept uncontaminated. If something isn't quite right with one testing procedure, the scientist can run another sample just to make sure. If the scientist runs out of blood, she can go back to the source and get more. By contrast, in the forensic setting, scientists often have only minute samples of blood, perhaps quite old, possibly contaminated with bacteria or other materials from the crime scene—and some of these contaminants might themselves contain DNA from other sources. Depending on the sample size and the techniques used, running one sample might use up all of the available blood—in which case, if anything goes wrong, the examiner would be out of luck. As Judge Sheindlin put it in his opinion in *Castro*, "for forensic purposes, there is only one bite at the apple."[14]

One of the ways that scientists using the new technology made sure that everything was working properly was by having a "control lane," in which they tested DNA from a known source, to make certain that all of the probes were working properly. In *Castro*, Lifecodes had properly used a control lane—but figuring out whose DNA had been used in the control lane became a comedy of errors. At first, Baird testified that the blood used in the control lane came from the HeLa cell line, a commercially available cell line widely-used in research and experimentation. But another Lifecodes employee testified that the blood probably came from a male scientist who worked at Lifecodes, and Baird subsequently agreed with this assessment. However, the control DNA had not reacted with a probe that targeted parts of the Y-chromosome—which, if the control DNA were male, it should have done. Baird explained that this Lifecodes employee must have an unusual genetic condition in which his Y chromosome was "short," and thus happened not to react with this probe.

Lander was skeptical. A genetic condition like the one Baird described would be extremely rare—and it would be both odd and extremely poor judgment to use for control purposes the DNA of an employee who was so genetically atypical. Upon further investigation, Lifecodes established that, sure enough, the "control" DNA belonged to a *different* Lifecodes employee, this time a female.

The defense suggested that the dreadful recordkeeping and sloppiness illustrated by Lifecodes' inability accurately to identify the source of the control DNA was both inexcusable and illustrative of a part of a more general pattern of unjustifiably poor quality control that made their results uncertain and untrustworthy. Lander detailed many other problems with the laboratory's records: failures to note experiments that had been run, failure to note dates accurately, failure to label autorads correctly. While granting that "no one of these things is fatal," Lander opined that "so many of them are questionable here that it makes me worry a great deal about whether a recognized procedure was in place for doing [these experiments]." In addition, it appeared that Lifecodes had knowingly used a probe they realized was contaminated. At one point on the autorad, there was an extra band on Ponce's DNA not visible on the blood from the watch. Lifecodes went to great length, and performed several experiments, to establish that this extra band was bacterial in origin—the result of a contaminated probe, rather than the result of an actual difference between Ponce's DNA and the bloodstain from the watch. The defense suggested that to continue to use a contaminated probe was scientifically unacceptable. While Baird explained that they had no choice, because making a new probe would have entailed significant delays, the defense claimed this was another example of sloppiness, and an example of how commercial interests had gotten in the way of doing science properly.

Then, at another locus, the blood taken from the watch appeared to reveal two bands not present in Ponce's blood sample. Baird testified that he was confident that these bands were the result of contamination and should therefore be ignored. But the defense claimed that Baird had no scientific basis for his confidence, and strongly criticized the laboratory for failing to do further experiments that could have definitively established whether or not these bands were genuine.

Lander also emphasized the danger of examiner bias: the danger that people—even scientists—tend to see what they are looking for when they interpret an autorad. He said that in his lab at MIT, they sometimes joked about the risk that a scientist might "hallucinate a band" when they expected to see one, and "just as one hallucinates bands where one expects to see them, one tends to discount things where one does not expect to see them." Therefore conducting the follow-up tests necessary to check any interpretations that might have been colored by prior expectations was absolutely critical. In his opinion, Lifecodes routinely failed to do that.

Baird, in retrospect, thought that the problem was in part the daily reality of forensic science being evaluated from the lofty perspective of research scientists. He said in a 1994 interview, "The reality is that when you do a test on a forensic sample, it is what it is, and you have to interpret it. It isn't my fault if the sample is contaminated or mixed or shitty. …I'm just trying to interpret what's there."[15]

The defense also raised important issues in *Castro* about the the use of population genetics in the case; that is, the methods by which Lifecodes determined not the fact of a match but its statistical meaning. To figure out how often one would expect to find two DNA samples that matched at a given set of loci, a scientist must have information about how frequently each allele is found in the population, and must also know to combine the likelihood of each particular allele into one combined frequency statistic. (To oversimplify slightly, one key issue was whether databases divided into major racial subgroups (e.g., Caucasian, Black, Hispanic) could actually provide adequate information about the frequency with which particular alleles were distributed in a population. Some believed these databases to be sufficient, while others thought that because some ethnic subgroups might tend to intermarry—e.g., Irish-Americans might tend to marry other Irish-Americans at a rate greater than chance—these general databases might misstate the expected allele distribution in particular subpopulations.) In Lander's view, scientists had not yet developed adequate knowledge about these matters to calculate the statistical meaning of a match with complete confidence.

All in all, the defense was able to show persuasively that Lifecodes had exercised poor judgment, engaged in shoddy quality control practices, and that their conclusions could not simply be presumed accurate. As Judge Sheindlin wrote in his opinion, "In a piercing attack upon each molecule of evidence presented, the defense was successful in demonstrating to this court that the testing laboratory failed in its responsibility to perform the accepted scientific techniques and experiments in several major respects."

An Extraordinary Meeting

Shortly after Eric Lander had finished testifying for the defense, he and Richard Roberts, one of the key witnesses for the prosecution in the case, ran into each other at a scientific meeting on genome mapping in Cold Spring Harbor. Lander gave Roberts a copy of his written report about Lifecodes' DNA evidence in the case and suggested that Roberts would likely find it to be interesting reading. Roberts was certainly troubled by what he read. In fact, he was so troubled that he proposed that all of the expert witnesses in the case—both prosecution and defense witnesses—should gather to talk about the issues, scientist to scientist, no lawyers allowed. Although eight of the ten witnesses who were contacted liked the idea of meeting, only four were able to fit the meeting into their schedules. On May 11, 1989—before the conclusion of the preliminary hearing—Lander, Roberts and two other witnesses (one from the defense, one from the prosecution) convened a mini-conference to see what they all thought about the evidence in *Castro*. They found that there was, indeed, much upon which they could agree.

After the meeting, the attendees issued a joint statement that left the prosecutors in *Castro* almost helpless: "Overall, the DNA data in this case are not scientifically reliable enough to support the assertion that the samples . . . do or do not match," they concluded. "If this data were submitted to a peer reviewed journal in support of a conclusion, it would not be accepted." The consensus of the experts on both sides that the evidence was invalid made it very difficult to imagine that the court would find it nonetheless to be "generally accepted" by the scientific community. All of the experts on both sides of the case, except for Michael Baird from Lifecodes, eventually endorsed the conclusions reached in this meeting.

The prosecution successfully kept the joint statement itself from being introduced in the hearing, on the grounds that it was hearsay. But the defense responded by calling several prosecution expert witnesses to the stand, who repeated under oath the conclusions that they had reached about the inadequacy of the particular DNA evidence in the case. In a way, the joint meeting and its consequences made Sheindlin's job in the *Castro* case a good deal easier: when the standard is "general acceptance" and the enormously-credentialed, hand-picked experts *on both sides* actually reach a *consensus*, who is the judge to second guess these shared conclusions of the experts?

Such a gathering of witnesses on both sides, *ex parte*, was of course highly irregular; with a flair for the dramatic, Peter Neufeld later called it "unprecedented in the annals of law." "We wanted to be able to settle the scientific issues through reasoned argument, to look at the evidence as scientists, not as adversaries," Richard Roberts explained afterwards to the press. "We all did so much better when we sat down without the lawyers, and had a reasoned scientific discussion. Perhaps it's time the system changed." Indeed, the scientists' joint statement criticized the use of the courtroom as a venue for reaching scientific consensus:

> All experts have agreed that the *Frye* test and the setting of the adversary system may not [be] the most appropriate method for reaching scientific consensus. The *Frye* hearing is not the appropriate time to begin the process of peer review of the data. . . . The setting also discourages many experts from agreeing to participate in the careful review of the data.[16]

The joint statement also called on the National Academy of Sciences to organize a committee to study the questions surrounding the use of forensic DNA.

Roberts, in particular, had harsh words for the adversarial process in his comments to the press. "Lawyers are more interested in getting certain words down on the written record than in arriving at the truth. Lawyers hope or want witnesses to say slightly more than they feel comfortable saying. . . . I do not find that the best way to reach the truth." While the scientists' discomfort with adversarial processes is understandable, it is also undoubtedly the case that *Castro*—the hearing itself, the ruling, and the significant publicity—revealed far more about how Lifecodes was conducting its DNA tests than any non-adversarial process had yet done. Outside of the setting of the courtroom, it is very difficult to imagine a research scientist of Eric Lander's caliber spending as much time and effort analyzing in detail the work product of a commercial forensic laboratory. The adversarial process has both flaws and excesses, but it also is a setting in which participants can drill down, analyze and unpack weaknesses in evidence in ways that may sometimes risk being unfair—but that also can be very revealing. Lander, for example, was able to work with Lifecodes' data and examine their protocols because courts can force the disclosure of material that in other settings could be kept confidential as trade secrets and proprietary information.

Roberts himself had testified in a number of earlier cases for Lifecodes, without ever having seen the kinds of data that Lander, Scheck and Neufeld had insisted upon in *Castro*. He simply assumed that the prosecutor and Lifecodes were showing him all of the relevant information. He had seen his role primarily as providing background information about DNA in general, but of course, both his presence and the substance of his testimony served to shore up the legitimacy of the forensic use of DNA in particular. Roberts may have criticized the adversary process, but it was this same process that led direclty to the production of information that changed his mind about the DNA sample in *Castro*.

To be sure, the use of expert evidence within an adversarial legal system has obvious dangers: it encourages expert participants to make stronger statements than they might in other settings, to become partisans rather than fair-minded evaluators or to overstate minor errors or mistakes that may be an inevitable part of any human endeavor. But it is important to recognize that adversarial methods can have a productive dimension as well: until the hearing in *Castro*, no one had any idea that Lifecodes was not following its own procedures and protocols in a variety of meaningful ways. Up to that point, neither scientific conferences, nor publication, nor peer review, nor internal laboratory checks or audits had brought to light what the adversary process made quite visible: both the significant deficiencies in how Lifecodes had handled the DNA in the Castro case, and more generally, that there were a number of important, not fully resolved problems relating to the transfer of DNA techniques into the forensic setting.

The Ruling

At the end of the preliminary hearing, Judge Sheindlin had presided over the longest, most in-depth legal examination of DNA profiling that had ever taken place. He had listened to days on end of testimony at the cutting edge of science, substance that he acknowledged was often far outside of his comfort zone. As he told the defense counsel at one point during Lander's cross-examination, "I don't have any scientists who I can ask about these things; therefore, I stand up here—sit up here alone attempting, as best as I can struggle. I work on [these issues] after Court session until late in the evening so that I can understand it."

Judge Sheindlin explained in his ruling that he would be guided by a three-prong test for examining whether the prosecution's DNA evidence met the *Frye* standard of general acceptance:

> Prong I. Is there a theory, which is generally accepted in the scientific community, which supports the conclusion that DNA forensic testing can produce reliable results?
>
> Prong II. Are there techniques or experiments that currently exist that are capable of producing reliable results in DNA identification and which are generally accepted in the scientific community?
>
> Prong III. Did the testing laboratory perform the accepted scientific techniques in analyzing the forensic sample in this particular case?[17]

Sheindlin recognized that courts often viewed the *Frye* test as encompassing only the first two prongs, figuring that the third prong—the case-specific implementation of the general tests—went to the weight of the evidence rather than its admissibility. Whether or not the third prong was appropriately defined as part of the *Frye* test or as something separate from it, Sheindlin thought that it was a crucial focus for pre-trial assessment of DNA. "[G]iven the complexity of the DNA multistem identification tests and the powerful impact that they may have on the jury, passing muster under *Frye* alone is insufficient to place this type of evidence before a jury without a preliminary, critical examination of the actual testing procedures performed in a particular case," he explained.[18]

In his ruling, Scheindlin went through each prong in turn, one by one. He first provided an introductory primer to both DNA identification in general (Prong I), and the forensic use of DNA for determining identification (Prong II). Prong I was, he thought, quite unproblematic: "The evidence in this case clearly establishes unanimity among all the scientists and lawyers as well that DNA identification is capable of producing reliable results." Sheindlin also answered Prong II, the question of whether there were presently techniques for reliably making DNA identifications in the *forensic* context, in the affirmative.

But when it came to the third prong, whether the specific tests were adequately performed by the laboratory in analyzing the DNA sample, in this case, Sheindlin's answer was a resounding "no." He spent several pages describing the "major respects" in which Lifecodes "failed to conduct the necessary and scientifically accepted tests," such as their unacceptable use of an apparently contaminated probe, their failure to use adequate controls for sex typing, their failure to do further tests to assess the two extra bands seemingly visible in the watch sample and their failure to use the same standards for measuring the existence of a match and assessing its statistical probability. As a result of these many lapses, Sheindlin concluded that he would permit at trial any evidence of exclusion—that is, that two samples did not match—but he would exclude the evidence suggesting a match between the watch sample and Vilma Ponce's DNA. In other words, the prosecution would be able to offer evidence that the blood found on the watch did not belong to Castro, but they would not be able to say that it was almost certainly Ponce's. Because Sheindlin decided that the evidence of a DNA match would not be permitted at trial, he deemed it unnecessary to delve into the questions of population genetics, an area that had played a relatively small role in *Castro* but would assume much greater importance in subsequent cases.

This marked the conclusion of what "some have referred to as the most comprehensive and extensive legal examination of DNA forensic identification tests held to date in the United States."[19] Sheindlin's decision marked the very first time that any American judge had restricted the use of DNA evidence in court. In addition, his opinion made clear that when scrutinized carefully, DNA tests in actual practice might turn out to have serious flaws.

And yet, without a doubt, Sheindlin's framework cabined the defense victory by making the emphasis quite particularistic and local, emphasizing Lifecodes' sloppy examination of *this* DNA comparison rather than recognizing problems with their forensic DNA analyses more generally. Given the many embarrassing revelations that had emerged at trial, and considering that nearly all the experts had reached a consensus that this particular DNA test could not be validly interpreted, an opinion that rejected the DNA evidence in this case without formally casting any doubt on the forensic use of DNA more generally was about the best outcome that the prosecution could have wished for. In fact, in their final brief, the prosecution had acknowledged that the DNA evidence in this case was insufficiently reliable: "Here, the People believe that we have not met our burden of demonstrating by a preponderance of the evidence that the accepted scientific techniques were utilized in this case. The scientific evidence generated in this case, as a whole, is too ambiguous to be admissible in a criminal case."[20]

By the time the hearing was over, the prosecution was thus granting that this particular DNA match was unreliable, and was hoping that Sheindlin would nonetheless recognize the general validity of forensic DNA typing. In this sense, although they, of course, had been forced by the defense over the course of the hearing to back-pedal considerably, the prosecution got exactly what it had hoped for from Judge Sheindlin's ruling. By contrast, Scheck and Neufeld were sorely disappointed: from their point of view, there was little reason to believe that Lifecodes had been unusually careless or sloppy in their testing of the evidence in *Castro*. It seemed clear to them that the problems with Lifecodes' protocols and quality control were both systemic and widespread, rather than the result of atypical lapses in this particular case alone. They would no doubt have preferred Sheindlin to have reached a more general conclusion, something that would have clearly signaled that forensic DNA, while highly promising, was not yet ready for prime time. They would have liked him to have recognized that the problems with Lifecodes' analysis were so serious as to implicate his second as well as his third prong.

Though Sheindlin's unwillingness to make his criticisms in a more generalized way greatly frustrated Scheck and Neufeld, Sheindlin's analysis under Prong II was not a complete whitewash of Lifecodes in particular or forensic DNA more generally. The opinion did recognize the importance of inquiring into technology *transfer*. Unlike several of the earlier judges who had assessed the admissibility of DNA evidence, Sheindlin well understood that DNA identification techniques' validity in other contexts did not necessarily translate into reliability in the forensic context, where there might be particular problems arising from the sometimes miniscule amounts of available biological material, from possible contamination or deterioration of the sample and from more difficult problems of measurement interpretation. Even though Sheindlin did conclude that the presently available techniques were adequate for dealing with these special difficulties of forensic DNA testing, the opinion was significant for at least recognizing them as difficulties that had to be dealt with.

Moreover, Sheindlin explicitly took issue with the widespread assumption, captured, for example by the court's ruling in *People v. Wesley*, that DNA testing would necessarily produce either the right answer or no answer at all. Sheindlin explained that while several earlier cases had suggested that "improper procedures and experiments will automatically and clearly be revealed, this court, on the contrary, advises caution in reviewing the procedures. For example, contaminated samples, probes or controls, may produce extra bands on the autorads which can cause differing scientific opinions in the interpretation of the autorads. On the other hand, degradation of a sample may fail to produce a band, again resulting in interpretation problems." Any court that took *Castro* seriously could no longer repeat the oft-made, comforting claim that there was no such thing as a false positive, nor buy into the implicit corollary that a DNA test was virtually self-interpreting.

Thus the case was quite a mixed result. Certainly it was a partial and significant victory for the defense, but at the same time, because it was so narrowly drawn, the Bronx district attorney's office could simultaneously call it a "victory of national importance" that reaffirmed the general validity and admissibility of DNA evidence. Interestingly, the opinion itself makes only a passing and opaque reference to the important fact that by the time the hearing had concluded, almost all of the experts for both sides (and even the prosecution itself) had conceded that the DNA test in this case was inconclusive. The savvy reader can find, in footnote 12, an aside mention that two of the prosecution's experts were recalled by the defense and, having earlier testified to the reliability of DNA identification, now allowed that the laboratory's lapses made this particular result inconclusive.

When Joseph Castro pleaded guilty to second degree murder on September 15, 1989, he admitted that the blood on the watch was likely that of Vilma Ponce after all. With that, the *Castro* case officially came to an end, but the controversies over DNA most certainly did not.

The Aftermath and the Broader Consequences of *Castro*

In the wake of *Castro*, Lifecodes made several significant changes to their internal procedures. For example, they began to use a computer-based matching system instead of relying only on visual comparison to declare a match, and they modified the way that they determined the frequency of alleles in their population databases: essentially, they took a number of Eric Lander's suggestions. (Lander himself was invited to testify in 57 DNA cases in the six months after *Castro*. Though he provided some technical assistance in a handful of select cases, he turned down all 57 of the offers to testify.)

In addition, the *Castro* hearing, along with the joint statement signed by the experts from both sides, fueled a growing belief that forensic DNA needed to be examined and studied by an authoritative, neutral group of scientists and other experts. In December 1989, the National Academy of Sciences appointed a committee to investigate and, if possible, forge a consensus, about the scientific resolution of the many technical and procedural issues surrounding the forensic use of DNA that the *Castro* hearing had highlighted.

Considering the many revelations of the pre-trial hearing, Sheindlin's opinion was about as narrowly-drawn as possible, but it was still a watershed moment for the forensic use of DNA. Along the way, the case had received a good deal of publicity, and newspapers in the months after *Castro* wrote about DNA quite differently than they had before. Doubt and uncertainty replaced the earlier tendency toward breathless enthusiasm. "DNA fingerprinting doesn't live up to initial promise," read one headline in the fall of 1989; "DNA 'Fingerprinting' Questioned; Geneticist Says Test May Be Less Reliable Than First Believed," said another. "Caution Urged on DNA Fingerprinting," warned *Science* magazine. "DNA Tests Unravel?" asked the *National Law Journal*. After *Castro*, journalists, the public, judges and jurors all became more willing to question DNA: it no longer seemed like an infallible magic bullet.

Castro affected the scientific landscape as well. In June, 1989, Eric Lander published an article in *Nature* concluding that the courts had been "too hasty" to accept DNA. He described in detail why forensic DNA fingerprinting is far more technically challenging than the diagnostic use of DNA. He also laid out a challenge to the scientific community: “It is my belief that we, the scientific community, have failed to set rigorous standards to which courts, attorneys, and forensic-testing laboratories can look for guidance—with the result that some of the conclusions presented to courts are quite unreliable.” Lander called in strong terms for both additional scientific study and greater regulation and oversight.[21] The case thus spurred greater scientific interest in the actual practices of forensic DNA testing and led to increased attention to questions of quality control, autorad interpretation and population genetics.

In this changed climate, defense attorneys became both more aggressive about challenging DNA and better able to locate the people and resources to mount effective challenges. In addition, more scientists began to evince a professional interest in the issues raised by forensic DNA evidence, especially in the questions surrounding population genetics and the statistical meaning of a match. After *Castro*, there is no doubt that DNA evidence in court received substantially more scrutiny—and a number of courts, including several state Supreme Courts, subsequently decided that problems with the DNA evidence made either restriction or exclusion necessary.

Indeed, over the next few years, the legal controversies over DNA increased in intensity and vociferousness. A *Los Angeles Times Magazine* article could write, in 1992, that “the battle over DNA fingerprinting has become the most entertaining and bewildering legal spectacle around.” While quality control issues remained significant, and questions of autorad interpretation received increased focus and attention, the most significant issue of all—both in the courtroom and in the pages of prestigious scientific journals—surrounded the questions of population genetics that Scheindlin’s opinion had punted. These battles raged not only in court but in prestigious scientific journals. For example, in a highly unusual move, the authors of an article on population genetics issues in the prestigious journal *Science* were asked by the editor to “tone down” their article, and the magazine—in part because of a recommendation by a member of their board of reviewing directors who also had a licensing relationship with Cellmark, one of the forensic DNA companies—decided to publish a simultaneous “rebuttal” alongside the original article.

In the meantime, the National Research Council (the research arm of the National Academy of Sciences) issued its long-awaited report in 1992. Instead of resolving disputes, the report generated new ones. It had proposed a compromise approach to the issue of population genetics that critics deemed scientifically unjustifiable, viewing it as an overtly political attempt to forge a compromise that lacked scientific foundation. In 1994 (and, not accidentally, just before the O.J. trial was beginning), the growing scientific consensus on these issues led Eric Lander and FBI DNA expert Bruce Budowle to publish a joint article in *Nature* entitled "DNA Fingerprinting Dispute Laid to Rest." While the O.J. Simpson case revealed starkly that DNA evidence was still controversial, the intense battles over admissibility of the technique itself largely came to an end. When the NRC issued a follow-up report in 1996, its recommendations were received with far less controversy.

A Precursor to *Daubert*

The significance of the *Castro* case goes beyond DNA itself in two important respects. First, the sheer detail and length of the hearing, and the tremendous focus on reliability—both of the technique in general and its particular application in the case—revealed a quite different approach to the evaluation of science in court than was typically seen under the *Frye* standard of general acceptance. *Castro* was an example of a growing trend by the courts to engage in the substantive assessment of the reliability of expert evidence, a trend that has only grown over the years since *Castro* was decided. In 1993, the Supreme Court decided in *Daubert v. Merrell Dow* that the Federal Rules of Evidence did not incorporate the *Frye* test of general acceptance, but that trial courts nonetheless had an obligation to serve as a gatekeeper with respect to expert evidence, to assure that it was sufficiently valid and reliable.[22] Although many states (including New York) have continued to use the *Frye* test, there has been an undeniable, though uneven, trend to examine expert evidence with increased scrutiny. Whether courts should be in the business of assessing the substance of scientific evidence—whether they have the know-how or the institutional competence—are certainly fair questions. But *Castro* is of a piece with this more general trend over the last several decades to examine scientific evidence proffered in court with increasing detail and care.

Castro as a Beacon: The Forensic Science Issue

If *Castro* stands for anything, it stands for the idea that the standards of research science are highly relevant for evaluating forensic science. Eric Lander's critique could be boiled down, in significant part, to the concern that Lifecodes was not taking issues of quality control, interpretation, and population genetics as seriously as an academic research laboratory would, and that given the stakes involved, this failure was unjustifiable and inexcusable. To meet the standards of academic scientific laboratories does not require perfection—time and time again, in his testimony, Lander emphasized that no laboratory operates completely without errors. But he saw no reason why commercially-run forensic science laboratories should be given anything approaching a free pass.

The DNA cases like *Castro*, along with *Daubert*'s increasing focus on judicial gatekeeping, have in recent years given ammunition to critics of many other forms of forensic science. Although some forensic science techniques have been in use for a century or more, many approaches to identification science, including handwriting identification evidence and fingerprinting, simply do not have the kind of empirical basis for their claims to validity that one would ordinarily associate with research science. These and similar forensic "sciences" may usually provide "right" answers—but because they have been subject to little rigorous validity testing, it is difficult to assess the real-world frequency of error. Of course, in *Castro* itself, notwithstanding the significant problems with Lifecodes' procedures, the laboratory's bottom-line conclusion that the watch stain matched Ponce's blood was, as far as anyone knows, correct—but Sheindlin's decision to exclude the evidence was nonetheless indisputably the right answer based on the record before him. Other kinds of forensic science evidence raise problems analogous to those faced in the *Castro* case. Although fingerprint experts testify that they can identify a match with 100 percent certainty and to the exclusion of all other fingerprints in the world, fingerprinting lacks any kind of statistical foundation. Just as Lifecodes' technicians' eyeballed an autorad to determine whether there was a match, fingerprint experts do not have formal standards or protocols for deciding when to declare a match. Nor do we have any real idea of how often, in the real world, fingerprint experts or handwriting identification experts might make honest mistakes in their evaluations. Fingeprint experts' frequent insistence that the technique is error free is reminiscent of the early—and erroneous—claims with DNA that there was no such thing as a false positive.[23]

The DNA cases, combined with *Daubert*, have led a set of defense lawyers to mount in recent years a number of challenges to other forms of forensic science. Some of the challenges to handwriting identification evidence have been successful; the challenges to fingerprinting largely have not. But *Castro* invites the question: should we evaluate forensic science differently from other kinds of scientific enterprises, and, if so, upon what justification? If not, then should these kinds of evidence be excluded or limited until further research and study validates both the proficiency of the examiners and the scientific bases for their claims? And so, we end with an irony: the technique that drew its earliest authority from a metaphoric association with "fingerprinting" may, in the end, help to reveal the weaknesses of fingerprinting and other forms of forensic science.

Endnotes

†This is an edited version of an essay originally published in EVIDENCE STORIES (Richard Lempert ed., 2006). Note that because the audience for the original essays was primarily advanced law students rather than scholars, an editorial decision was made within the volume as a whole to keep footnotes to a minimum, citing only key sources and direct quotations.

1 This description of the facts draws extensively on the detailed account in HARLAN LEVY, AND THE BLOOD CRIED OUT: A PROSECUTOR'S SPELLBINDING ACCOUNT OF THE POWER OF DNA (1996). It is based as well on Roger Parloff*, How Barry Scheck and Peter Neufeld Tripped up the DNA Experts*, THE AMERICAN LAWYER, Dec. 1989, at 50; Edward Humes, *The DNA Wars; Touted as an Infallible Method to Identify Criminals, DNA Matching has Mired Courts in a Vicious Battle of Expert Witnesses*, L.A. TIMES, Nov. 29, 1992, (Magazine) at 20; HOWARD COLEMAN & ERIC SWENSON, DNA IN THE COURTROOM: A TRIAL WATCHER'S GUIDE (1994).

2 Alec Jeffreys, interview with Michael Lynch, 6 August 1996, cited in Jay Aronson, The Introduction, Contestation and Regulation of Forensic DNA Analysis in the American Legal System (1984-1995), 56 (Ph.D. Dissertation, Univ. of Minnesota, 2004). Aronson's dissertation, a detailed and insightful account of the rise of DNA as a form of legal evidence, and, apart from the *Castro* trial transcript, the most important source for this essay, will be published in book form as THE DEVELOPMENT OF DNA PROFILING: SCIENCE, LAW & CONTROVERSY IN THE AMERICAN JUSTICE SYSTEM (Rutgers University Press, forthcoming 2007)

3 This account of the first legal use forensic DNA is drawn from ARONSON, *supra* note 2 as well as contemporaneous newspaper articles.

4 For a detailed and novelistic book-length account of this case, upon which this short summary is based, see JOSEPH WAMBAUGH, THE BLOODING (1989).

5 Although a detailed explanation of the differences between Lifecodes and Cellmark's early techniques is beyond the scope of this essay, here is a quick explanation: Jeffreys and Cellmark's technique originally used a 'multi-locus probe' (MLP) that bound to many loci in a person's DNA and produced an image that looked like a complex pattern, while Lifecodes approach was to use 'single locus probes." Individually, these single locus probes could not provide as much information as a MLP, but they could be aggregated to build up information about an individual's genetic profile and they could be used on smaller amounts of blood, and were easier to interpret than MLPs. Over time, SLP's came to be the dominant approach.

6 Andrews v. State, 533 So. 2d 841 (Fla. App. 1988).

7 Aronson, *supra* note 2 at 128.

8 140 Misc. 2d 306, 308-09 (N.Y. County Ct. 1988).

9 Frye v. United States, 293 F. 13 (D.C. App. 1923).

10 *Id.* The court did give a certain grudging credence to arguments the defense had made about the inadequate database-size used for the population genetics, and thus required the prosecution to reduce the probabilities by a factor of 10, permitting them at trial to claim identification ability at the level of 1 in 84,000,000 for American Caucasians and 1 in 140,000,000 for American Blacks.

11 Parloff, *supra* note 1.

12 A copy of the hearing transciript was obtained from the O.J. Simpson Archive at Cornell University. Other important sources on the hearing include ARONSON, *supra* note 2, Parloff, *supra* note 1, and, the numerous newspaper accounts written during and after the hearing. Quotations in this section come from the trial transcript unless otherwise noted.

13 Michael Baird, interview with Saul Halfon and Arthur Daemmerich, 14 July, 1994 (O.J. Simpson Murer Trial and DNA Typing Archive, Cornell University, #53/12/3037, Box 2, Division of Rare and Manuscripts Collections, Cornell University Library), *quoted in* ARONSON, *supra* note 2 at 221.

14 People v. Castro, 144 Misc. 2d 956, 970 (NY 1989). DNA technology has, however, changed since *Castro.* Techniques for multiplying minute quantities of DNA now allow DNA comparisons that were impossible in 1989.

15 Michael Baird interview, *supra* note 13.

16 "Statement of the Independent Expert Scientists Having Testified in the Frye Hearing in People v. Castro," 11 May 1989 (copy on file with author).

17 People v. Castro, 144 Misc. 2d, 956, 960 (1989).

18 Castro, 144 Misc. 2d 956.

19 Castro, 144 Misc. 2d at 960.

20 Timothy Clifford*, DNA-Test Errors Conceded*, NEWSDAY (July 4, 1989 at p. 7) (quoting from memorandum submitted by the prosecution in *People v. Castro*.)

21 Lander noted that "at present, forensic science is virtually unregulated—with the paradoxical result that clinical laboratories must meet higher standards to be allowed to diagnose strep throat than forensic labs must meet to put a defendant on death row." Eric Lander, *DNA Fingerprinting on Trial*, 339 NATURE 501, 505 (1989).

22 509 U.S. 579 (1993).

23 An historian of fingerprinting has recently documented 22 publicly-known instances of fingerprinting identification error, and argues persuasively that these known misattributions probably account for only a small fraction of the mistakes that have actually been made. *See* Simon Cole, *More Than Zero. Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985 (2005).