The Coming Paradigm Shift in Forensic Identification Science

Michael J. Saks[1] and Jonathan J. Koehler[2]

Converging legal and scientific forces are pushing the traditional forensic identification sciences toward fundamental change. The assumption of discernible uniqueness that resides at the core of these fields is weakened by evidence of errors in proficiency testing and in actual cases. Changes in the law pertaining to the admissibility of expert evidence in court, together with the emergence of DNA typing as a model for a scientifically defensible approach to questions of shared identity, are driving the older forensic sciences toward a new scientific paradigm.

Little more than a decade ago, forensic individualization scientists compared pairs of marks (handwriting, fingerprints, tool marks, hair, tire marks, bite marks, etc.), intuited whether the marks matched, and testified in court that whoever or whatever made one made the other. Courts almost never excluded the testimony. Cross-examination rarely questioned the foundations of the asserted expertise or the basis of the analyst's certainty.

Today, that once-complacent corner of the law and science interface has begun to unravel—or at least to regroup. The news carries reports of erroneous forensic identifications of hair, bullets, handwriting, footprints, bite marks, and even venerated fingerprints. Scientists have begun to question the core assumptions of numerous forensic sciences (*1*–*6*). Federal funding has materialized to support research that examines long-asserted but unproven claims. Courts have started taking challenges to asserted forensic science expertise seriously (*1*). A dispassionate scientist or judge reviewing the current state of the traditional forensic sciences would likely regard their claims as plausible, underresearched, and oversold.

The traditional forensic individualization sciences rest on a central assumption: that two indistinguishable marks must have been produced by a single object. Traditional forensic scientists seek to link crime scene evidence to a single person or object "to the exclusion of all others in the world" (*7*, *8*). They do so by leaning on the assumption of discernible uniqueness. According to this assumption, markings produced by different people or objects are observably different. Thus, when a pair of markings is not observably different, criminalists conclude that the marks were made by the same person or object.

Although lacking theoretical or empirical foundations, the assumption of discernible uniqueness offers important practical benefits to the traditional forensic sciences. It enables forensic scientists to draw bold, definitive conclusions that can make or break cases. It excuses the forensic sciences from developing measures of object attributes, collecting population data on the frequencies of variations in those attributes, testing attribute independence, or calculating and explaining the probability that different objects share a common set of observable attributes. Without the discernible uniqueness assumption, far more scientific work would be needed, and criminalists would need to offer more tempered opinions in court.

Legal and scientific forces are converging to drive an emerging skepticism about the claims of the traditional forensic individualization sciences. As a result, these sciences are moving toward a new scientific paradigm. [We use the notion of paradigm shift not as a literal application of Thomas Kuhn's concept (*9*), but as a metaphor highlighting the transformation involved in moving from a prescience to an empirically grounded science.] Two such forces are outgrowths of DNA typing: the discovery of erroneous convictions and a model for a scientifically sound identification science. A third force is the momentous change in the legal admissibility standards for expert testimony. A final force grows from studies of error rates across the forensic sciences.

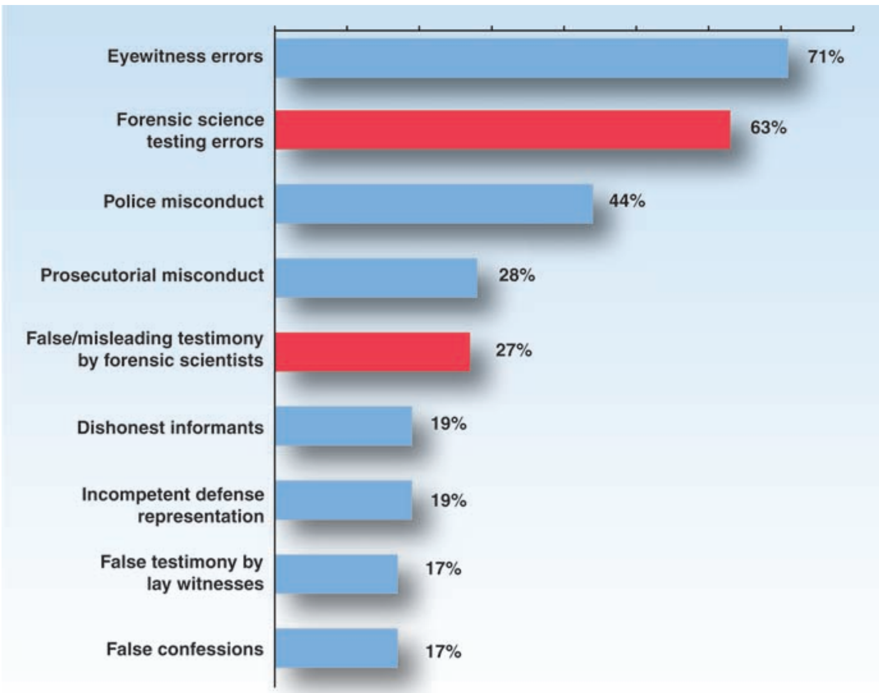

Fig. 1. Factors associated with wrongful conviction in 86 DNA exoneration cases, based on case analysis data provided by the Innocence Project, Cardozo School of Law (New York, NY), and computed by us. Percentages exceed 100% because more than one factor was found in many cases. Red bars indicate factors related to forensic science.

[1]College of Law, Arizona State University, Tempe, AZ 85287, USA. E-mail: saks@asu.edu [2]McCombs School of Business, University of Texas, Austin, TX 78712, USA. E-mail: koehler@mail.utexas.edu

Fig. 2. Bite mark evidence exhibit from trial of Ray Krone, suggesting alignment of a cast of Krone's dentition with bite wounds in victim's flesh [*State v. Krone*, 182 Ariz. 319 (1995)]. A forensic odontologist testified that this showed Krone to be the biter. Krone was convicted of murder and sentenced to death, but a decade later he was exonerated by DNA analysis. [Source: E. Thomas Barham (Los Alamitos, CA) and Alan Simpson (Phoenix, AZ), attorneys for Krone]

Post-Conviction DNA Exonerations

During the past decade, scores of people who were convicted of serious crimes—including at least 14 who had been sentenced to death—have been exonerated by DNA analyses of crime scene evidence that had not been tested at the time of their trials (*10*). It was not surprising to learn that erroneous convictions sometimes occur, and that new science and technology can help detect and correct those mistakes. Nor was it surprising to learn, from an analysis of 86 such cases (Fig. 1), that erroneous eyewitness identifications are the most common contributing factor to wrongful convictions. What was unexpected is that erroneous forensic science expert testimony is the second most common contributing factor to wrongful convictions, found in 63% of those cases. These data likely understate the relative contribution of forensic science expert testimony to erroneous convictions. Whereas lawyers, police, and lay witnesses participate in virtually every criminal case, forensic science experts participate in a smaller subset of cases—about 10 to 20% of criminal cases during the era when these DNA exonerations were originally tried (*11*).

Figure 1 also indicates that forensic scientists are the witnesses most likely to present misleading or fraudulent testimony. Deceitful forensic scientists are a minor sidelight to this paper, but a sidelight that underscores cultural differences between normal science and forensic science (*12*, *13*). In normal science, academically gifted students receive four or more years of doctoral training where much of the socialization into the culture of science takes place. This culture emphasizes methodological rigor, openness, and cautious interpretation of data. In forensic science, 96% of positions are held by persons with bachelor's degrees (or less), 3% master's degrees, and 1% Ph.D.s (*14*). When individuals who are not steeped in the culture of science work in an adversarial, crime-fighting culture, there is a substantial risk that a different set of norms will prevail. As one former forensic scientist noted, this pressure-packed environment can lead to data fudging and fabrication: "All [forensic science] experts are tempted, many times in their careers, to report positive results when their inquiries come up inconclusive, or indeed to report a negative result as positive" [(*15*), p. 17].

DNA Typing as the New Model for Scientific Forensic Identification

Much of the above criticism does not apply to the science of DNA typing as practiced today. Indeed, DNA typing can serve as a model for the traditional forensic sciences in three important respects. First, DNA typing technology was an application of knowledge derived from core scientific disciplines. This provided a stable structure for future empirical work on the technology. Second, the courts and scientists scrutinized applications of the technology in individual cases. As a result, early, unscientific practices were rooted out. Third, DNA typing offered data-based, probabilistic assessments of the meaning of evidentiary "matches." This practice represented an advance over potentially misleading match/no-match claims associated with other forensic identification sciences.

Immediately after DNA's first courtroom appearance in the 1980s, scientists from disciplines as varied as statistics, psychology, and evolutionary biology debated the strengths and limitations of forensic DNA evidence. Blue-ribbon panels were convened, conferences were held, unscientific practices were identified, data were collected, critical papers were written, and standards were developed and implemented. The scientific debates focused on the adequacy of DNA databases (*16*), the computation of DNA match probabilities (*17*), the training of DNA analysts (*18*), the presentation of DNA matches in the courtroom (*19*), and the role of error rates (*20*). In some cases, disputants worked together to find common ground (*21*). These matters were not resolved by the forensic scientists themselves, by fiat, or by neglect. Most exaggerated claims and counterclaims about DNA evidence have been replaced by scientifically defensible propositions. Although some disagreement remains (*22*), the scientific process worked.

One of the great strengths of DNA typing is that it uses a statistical approach based on population genetics theory and empirical testing. Experts evaluate matches between suspects and crime scene DNA evidence in terms of the probability of random matches across different reference populations (e.g., different ethnicities). These probabilities are derived from databases that identify the frequency with which various alleles occur at different locations on the DNA strand. The traditional forensic sciences could and should emulate this approach (*23*). Each subfield must construct databases of sample characteristics and use these databases to support a probabilistic approach to identification. Fingerprinting could be one of the first areas to make the transition to this approach because large fingerprint databases already exist. The greatest challenge in this effort would be to develop measures of the complex images presented by fingerprints, tool marks, bite marks, handwriting, etc. (Figs. 2 and 3). Forensic scientists will need to work with experts in differential geometry, topology, or other fields to develop workable measures.

A second data collection effort that would strengthen the scientific foundation of the forensic sciences involves estimating error rates. Although the theoretical promise of forensic technology is considerable, the practical value of any particular technology is limited by the extent to which potentially important errors arise. The best way to identify the frequency with which errors occur is to conduct blind, external proficiency tests using realistic samples. A proficiency test requires analysts to make judgments about samples whose properties are known. External proficiency tests are conducted by an agency unaffiliated with the forensic scientist's laboratory. Externality is important to the integrity of proficiency tests because laboratories have strong incentives to

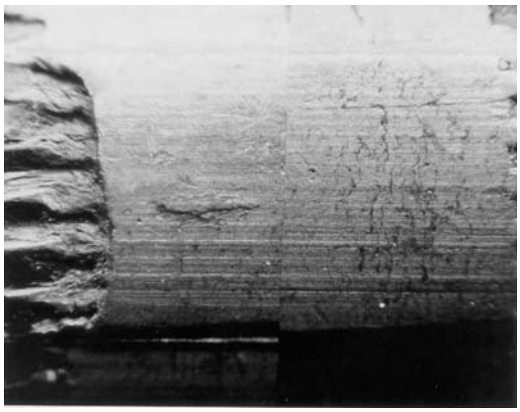

Fig. 3. Image of two bullets viewed through a comparison microscope. The bullets were fired from two consecutively manufactured Smith & Wesson 38 Special revolver barrels. Whether fired through the same or different barrels, numerous matching and nonmatching striations are engraved onto bullets. To reliably identify the barrel through which a questioned bullet was fired, an examiner must distinguish among class, subclass, and individual characteristics. These two bullets illustrate subclass characteristic agreement of striated markings on a groove impression that could be mistaken for individual characteristics. Without investigating the potential for subclass carryover, the examiner could mistake these as having been fired from the same gun. [Source: Bruce Moran, firearms examiner with the Sacramento County (CA) District Attorney, Laboratory of Forensic Services]

be perceived as error-free. An even better test would be a blind proficiency test, in which the analyst believes the test materials are part of ordinary case work. Blindness increases the validity of proficiency test results because it ensures that analysts treat the test sample as they would other case samples. Although proficiency tests are used in many forensic sciences, the tests are generally infrequent, internal, and unrealistic; blind tests are practically nonexistent.

Changes in the Law

Until recently, courts assessed expertise by looking for superficial indicia of validity. In the 19th century, courts were impressed by "qualifications" and success in the marketplace. If the market valued an asserted expertise or expert, courts generally did, too. In *Frye v. United States* [293 F. 1013 (D.C. Cir. 1923)], a federal appellate court confronted the question of admissibility of an expertise that had no life in any commercial marketplace. The court solved the problem by substituting an intellectual marketplace. The court asked whether the proffered expertise had "gained general acceptance in the particular field in which it belongs." Sixty years later, the *Frye* test had become the dominant expert evidence filter in American courts.

In 1993, the law began to catch up with the scientific method. In *Daubert v. Merrell Dow Pharmaceuticals* [509 U.S. 579 (1993)], the U.S. Supreme Court introduced a new standard for the admissibility of scientific evidence. Under *Daubert*, proffered scientific testimony must be shown to stand on a dependable foundation. The court suggested that trial judges making this determination consider whether the proffered science has been tested, the methodological soundness of that testing, and the results of that testing. The *Daubert* test in effect lowers the threshold for admission of sound cutting-edge science and raises the threshold for long-asserted expertise that lacks a scientific foundation. Seriously applied, the *Daubert* test subjects the forensic sciences to a first-principles scientific scrutiny that poses a profound challenge to fields that lack rigorous supporting data.

United States v. Starzecpyzel [880 F. Supp. 1027 (S.D.N.Y. 1995)] offered an early indication of how *Daubert* could change judicial views. After an extensive hearing on the soundness of asserted handwriting identification expertise, a federal district court concluded that the field had no scientific basis: "[T]he testimony at the *Daubert* hearing firmly established that forensic document examination, despite the existence of a certification program, professional journals and other trappings of science, cannot, after *Daubert*, be regarded as 'scientific … knowledge'" (p. 1038). However, the court did not exclude this unscientific testimony. It reasoned that handwriting identification did not have to reach the *Daubert* standard because *Daubert* applied only to scientific evidence, and handwriting identification plainly was not scientific evidence. Thus, when a forensic science was found to stand on a weak foundation, the threshold of admission was lowered to accommodate this weakness.

In *Kumho Tire v. Carmichael* [526 U.S. 137 (1999)], the Supreme Court directly confronted the question of whether *Daubert* applies to nonsciences. A consortium of law enforcement organizations prepared an amicus brief urging that *Daubert* scrutiny not be extended to the testimony of police agency expert witnesses. The brief argued that "the great bulk of expert testimony provided by law enforcement officers does not involve scientific theories, methodologies, techniques, or data in any respect…. Instead, law enforcement officers testify about such things as accident reconstruction, fingerprint, footprint and handprint [identification], handwriting analysis, firearms markings and toolmarks and the unique characteristics of guns, bullets, and shell casings, and bloodstain pattern identification" (*24*). Ironically, then, fields that initially gained entry to the courts by declaring themselves to be "sciences" now sought to remain in court by denying any connection with scientific methods, data, or principles. Despite efforts to preserve the "nonscience" loophole, the Supreme Court doctrinally sealed it shut when *Kumho Tire* held that all expert testimony must pass appropriate tests of validity to be admissible in court.

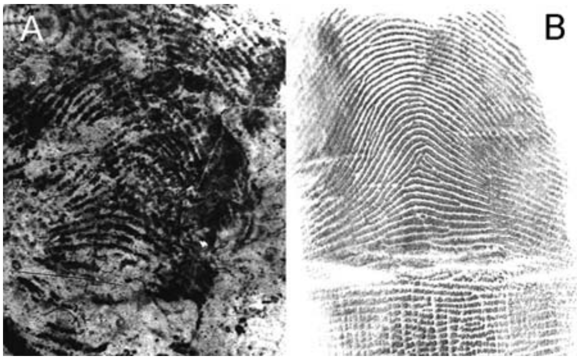

Fig. 4. (**A**) A latent fingerprint believed to belong to a terrorist involved in train bombings in Madrid, Spain, in March 2004. (**B**) A database print belonging to Brandon Mayfield of Portland, Oregon. On the basis of these prints (though not necessarily these very images), FBI fingerprint examiners erroneously identified Mayfield as the bomber (*26*). [Source: Problem Idents, onin.com/fp/problemidents.html#madrid]

Error Rates

Although *Daubert*'s testing recommendations are familiar to most scientists, there has been remarkably little research on the accuracy of traditional forensic sciences. Proficiency tests in some fields offer a step in the right direction, even though simple tasks and infrequent peer review limit their value. Nonetheless, the available data hint that some forensic sciences are best interpreted in tandem with error rates estimated from sound studies.

Unfortunately, forensic scientists often reject error rate estimates in favor of arguments that theirs is an error-free science. For example, an FBI document section chief asserted that all certified document examiners in the United States would agree with his conclusions in every case [(*25*), p. 196]. Likewise, fingerprint experts commonly claim that all fingerprint experts would reach the same conclusions about every print (*2*). Such hubris was on display in spring 2004 when the FBI declared that a fingerprint recovered from a suspicious plastic bag near the scene of a terrorist bombing in Madrid provided a "100 percent match" to an Oregon attorney (Fig. 4). The FBI eventually conceded error when Spanish fingerprint experts linked the print to someone else (*26*).

The FBI and other agencies often seek to preserve the illusion of perfection after disclosure of such errors by distinguishing between human errors ("possible") and errors of method ("impossible"). A leading FBI scientist explained the distinction to the court in *United States v. Llera-Plaza I* [58 Fed. R. Evid. Serv. 1 (E.D. Pa. 2002)]: "We have to understand that error rate is a difficult thing to calculate. I mean, people are trying to do this, it shouldn't be done, it can't be done…. An error rate is a wispy thing like smoke, it changes over time…. If you made a mistake in the past, certainly that's valid information … but to say there's an error rate that's definable would be a misrepresentation…. Now, error rate deals with people, you should have a method that is defined and stays within its limits, so it doesn't have error at all. So the method is one thing, people making mistakes is another issue."

Such claims are problematic. First, the suggestion that humans err but forensic techniques do not is unfalsifiable. It is impossible to disentangle "method" errors from "practitioner" errors in fields where the method is primarily the judgment of the examiner. Second, even if such disentanglement were possible, it is a red herring. When fact-finders hear evidence of a forensic match, a proper assessment of the probative value of that match requires awareness of the chance that a mistake was made. The source of such a mistake is irrelevant for this purpose. If method errors could be distinguished from practitioner errors, a 1% method error affects the probative value of the match in exactly the same way as a 1% practitioner error. Identifying sources of error is relevant for improving forensic science practice, but it plays no role in identifying the probative importance of a match.

Third, the suggestion that error rates do not exist because they change over time and are not specific to the case at hand is a base-rate fallacy. In this fallacy of reasoning, people underuse (or willfully ignore) general background data in judgment tasks because they believe the data are irrelevant to the instant case. However, general background data (or base rates) are relevant for specific predictions (*27*, *28*). For example, although risk estimates for a disease fluctuate and are developed on patients other than the patient now seeking medical advice, these estimates provide information useful for predicting whether this patient will contract the disease. A 20% base-rate risk of contracting the disease makes it more likely that the patient will get the disease than would a 1% risk. Likewise, an X% base-rate risk of error in a given forensic science provides some indication of the chance that a particular conclusion is in error (*22*).

Data from proficiency tests and other examinations suggest that forensic errors are not minor imperfections. Spectrographic voice identification error rates are as high as 63%, depending on the type of voice sample tested [(*1*), chap. 31]. Handwriting error rates average around 40% and sometimes approach 100% [(*1*), chap. 28]. False-positive error rates for bite marks run as high as 64% [(*1*), chap. 30]. Those for microscopic hair comparisons are about 12% (using results of mitochondrial DNA testing as the criterion) (*29*). Fingerprint examiners generally fare better, although data from a well-known forensic testing program contradict industry boasts of perfect, or even near-perfect, agreement (*30*). Since 1995, about one-fourth of examiners failed to correctly identify all latent prints in this test (which includes 9 to 12 latent prints and palmprints). About 4 to 5% of examiners committed false-positive errors on at least one latent. In one test, 20% of examiners mistook one person's prints for those of his twin. The editor of the leading fingerprint journal called this performance ''unacceptable'' [(*31*), p. 524]. It is noteworthy that these misidentifications are not confined to a single lab, circumstance, or marking. Moreover, the misidentification rates do not show a clear pattern of improvement (the misidentification rates in 2004 were 4 to 6%). Nor are these errors limited to arguably artificial testing situations; erroneous fingerprint identifications have made their way out of the crime lab and into prosecutions in at least 21 documented cases (*32*).

Forensic science proficiency tests and examinations are obviously imperfect indicators of the rate at which errors occur in practice. This fact does not justify ignoring the worrisome data these tests have yielded. Indeed, these data are probably best regarded as lower-bound estimates of error rates. Because the tests are relatively easy (according to test participants), and because participants know that mistakes will be identified and punished, test error rates (particularly the false-positive error rate) probably are lower than those in everyday casework (*33*, *34*).

The studies mentioned above cry out for attention and follow-up investigations. In light of the law's growing reluctance to accept experts' personal guarantees in lieu of scientific data, these studies should increase candor about performance and create pressure for improvement.

The Future

The traditional forensic sciences need look no further than their newest sister discipline, DNA typing, for guidance on how to put the science into forensic identification science. This effort should begin with adoption of the basic-research model. Just as DNA scientists tested the genetic assumptions that undergirded DNA typing theory (e.g., Hardy-Weinberg equilibrium), traditional forensic scientists should design experiments that test the core assumptions of their fields. As basic research knowledge grows, experts will be able to inform courts about the relative strengths and weaknesses of their theories and methods, and suggest how that knowledge applies to individual cases.

At the same time, data should be collected on the frequency with which markings and attribute variations occur in different populations. In addition to their case-specific benefits, these data may also facilitate the development of artificial intelligence or computer-aided pattern recognition programs for the identification sciences. Forensic scientists might also adopt protocols, such as blind examinations in combination with realistic samples, that minimize the risks that their success rates will be inflated and their conclusions biased by extraneous evidence and assumptions (*34*). When matches are identified, forensic scientists in all fields would compute and report random-match probabilities similar to those used in DNA typing. These estimates—in combination with error rate estimates provided by mandatory, well-constructed proficiency tests—would inform fact-finders about the probative value of the evidentiary match.

Simply put, we envision a paradigm shift in the traditional forensic identification sciences in which untested assumptions and semi-informed guesswork are replaced by a sound scientific foundation and justifiable protocols. Although obstacles exist both inside and outside forensic science, the time is ripe for the traditional forensic sciences to replace antiquated assumptions of uniqueness and perfection with a more defensible empirical and probabilistic foundation.

References

1. D. L. Faigman, D. Kaye, M. J. Saks, J. Sanders, *Modern Scientific Evidence: The Law and Science of Expert Testimony* (Thompson-West, St. Paul, MN, ed. 2, 2002).
2. R. Epstein, *S. Cal. Law Rev.* **75**, 605 (2002).
3. M. J. Saks, *Hastings Law J.* **49**, 1069 (1998).
4. National Research Council, *Forensic Analysis: Weighing Bullet Lead Evidence* (National Academy Press, Washington, DC, 2004).
5. I. Pretty, D. Sweet, *Sci. Justice* **41**, 85 (2001).
6. C. M. Cooley, *Stanford Law Policy Rev.* **15**, 381 (2004).
7. D. A. Stoney, *J. Forensic Sci. Soc.* **31**, 197 (1991).
8. J. W. Osterburg, *J. Crim. Law Criminol.* **60**, 97 (1969).
9. T. S. Kuhn, *The Structure of Scientific Revolutions* (Univ. of Chicago Press, Chicago, 1962).
10. B. Scheck, P. Neufeld, J. Dwyer, *Actual Innocence* (Doubleday, New York, 2000).
11. J. Peterson, Forensic Science and the Courts (National Institute of Justice, Washington, DC, January 1986).
12. P. C. Giannelli, *Issues Sci. Tech.* **20**, 55 (fall 2003).
13. P. C. Giannelli, *Va. J. Soc. Policy Law* **4**, 439 (1997).
14. K. G. Furton, Y. L. Hsu, M. D. Cole, *J. Forensic Sci.* **44**, 128 (1999).
15. A. A. Moenssens, *J. Crim. Law Criminol.* **84**, 1 (1993).
16. Committee on DNA Technology in Forensic Science, *DNA Technology in Forensic Science* (National Research Council, Washington, DC, 1992).
17. R. C. Lewontin, D. L. Hartl, *Science* **254**, 1745 (1991).
18. Technical Working Group on DNA Analysis Methods, *Crime Lab. Digest* **18**, 44 (1991).
19. K. Roeder, *Stat. Sci.* **9**, 263 (1994).
20. J. J. Koehler, A. Chia, J. S. Lindsey, *Jurimetrics J.* **35**, 201 (1995).
21. B. Budowle, E. S. Lander, *Nature* **371**, 735 (1994).
22. J. J. Koehler, *Jurimetrics J.* **37**, 425 (1997).
23. M. J. Saks, J. J. Koehler, *Cardozo Law Rev.* **13**, 361 (1991).
24. Brief Amicus Curiae of Americans for Effective Law Enforcement *et al.*, submitted in *Kumho Tire v. Carmichael*, No. 97-1709, in the Supreme Court of the United States (October Term, 1997).
25. D. Fisher, *Hard Evidence* (Simon & Schuster, New York, 1995).
26. D. Heath, H. Bernton, Portland lawyer released in probe of Spain bombings, *Seattle Times* (21 May 2004).
27. D. Kahneman, A. Tversky, *Psychol. Rev.* **80**, 237 (1973).
28. J. J. Koehler, *Behav. Brain Sci.* **19**, 1 (1996).
29. M. M. Houck, B. Budowle, *J. Forensic Sci.* **47**, 964 (2002).
30. Collaborative Testing Services Inc., Forensics Testing Program: Latent Prints Examination, 1993–2004.
31. D. Grieve, *J. Forensic Ident.* **46**, 524 (1996).
32. S. Cole, *J. Crim. Law Criminol.*, in press (available at www.seweb.uci.edu/faculty/cole).
33. J. L. Peterson *et al.*, *Forensic Evidence and the Police* (National Institute of Justice, Washington, DC, 1984).
34. M. J. Saks, D. M. Risinger, R. Rosenthal, W. Thompson, *Sci. Justice* **43**, 77 (2003).

10.1126/science.1111565