



ANNUAL
REVIEWS **Further**

Click [here](#) for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It

Michael J. Saks¹ and David L. Faigman²

¹College of Law and Department of Psychology, Arizona State University, Tempe, Arizona 85287; email: saks@asu.edu

²Hastings College of the Law, University of California, San Francisco, California 94102; email: faigmand@uchastings.edu

Annu. Rev. Law Soc. Sci. 2008. 4:149–71

First published online as a Review in Advance on July 8, 2008

The *Annual Review of Law and Social Science* is online at lawsocsci.annualreviews.org

This article's doi:
10.1146/annurev.lawsocsci.4.110707.172303

1550-3585/08/1201-0149\$20.00

Key Words

courts, *Daubert*, science, scientific evidence

Abstract

A group of nonscience forensic sciences has developed over the past century. These are fields within the broader forensic sciences that have little or no basis in actual science. They are not applications of established basic sciences, they have not systematically tested their own hypotheses, and they make unsupported assumptions and exaggerated claims. This review explains the nature and origins of those nonscience forensic fields, which include the forensic individualization sciences and certain other areas, such as fire and arson investigation. We explore the role of the courts in maintaining the underdeveloped state of these fields and consider suggestions for improving this state of affairs (addressing the potential role that could be played by these fields themselves, by the courts, and by normal sciences).

INTRODUCTION

Surprisingly little of what most people think they know about major aspects of the forensic sciences is as they imagine it to be. This review focuses on what a recent report of the National Institute of Justice (2007) refers to as the soft forensic sciences, or what might be more accurately described as the nonscience forensic sciences. The public appears to consider the various forensic sciences as one whole, believing them all to be flawless applications of sound basic science to problems of law. Courts generally share this view. As a Utah Court of Appeals judge commented in regard to fingerprint expert testimony: “In essence, we have adopted a cultural assumption that a government representative’s assertion that a defendant’s fingerprint was found at a crime scene is an infallible fact, and not merely the examiner’s opinion” (*State v. Quintana* 2004, Thorne, J., concurring, p. 171). The belief in fingerprint technology may be the most extreme example, but it is by no means the only. The trust that is laid upon the forensic sciences generally falls somewhere between uncritical faith and manufactured myth.

The nonscience forensic sciences, as the paradoxical phrase suggests, are those fields within forensic science that have little or no basis in actual science. They neither borrow from established science nor systematically test their hypotheses. Their primary claims for validity rest on anecdotal experience and proclamations of success over time. Hypothesis and supposition are typically considered sufficient. Whereas in most scientific fields experience and observation are designated as the first steps of the scientific method, for many forensic fields they constitute the final stages of confirmation. Indeed, in a way, many practitioners of the forensic arts have turned the scientific method on its head. So long as their hypotheses and suppositions have not been tested, they are assumed true. Hypotheses that endure over time, rather than actually being tested, are deemed proven. This model was once pervasive in applied settings, especially in medicine, and

produced such time-tested technologies as bloodletting and phrenology.

The fields that most resemble those ancient nonscience sciences are the forensic identification sciences, as well as certain other specialties within forensic science. By identification sciences we mean those subfields that often are referred to as criminalistics and that involve pattern matching in an effort to associate a crime scene mark or object with its source. These subfields include the comparison of fingerprints, handwriting, bitemarks, voiceprints, toolmarks, firearms, tire prints, shoe prints, and so on. Their goal is to link a latent fingerprint, a writing, a bitemark, a bullet, or similar objects to the one and only finger, writer, teeth, gun, or other specific object in the world that made the markings. The ultimate objective, and postulated achievement, of forensic identification science is individualization, “[t]he process of placing an object in a category which consists of a single, solitary unit. Individualization implies uniqueness . . .” (Thornton & Peterson 2008, p. 71). “Criminalistics is the science of individualization” (Osterburg 1969, p. 97). For example, a firearms examiner testifying in a federal court claimed to be able to identify the unknown weapon “to the exclusion of every other firearm in the world” (*U.S. v. Green* 2005, p. 107).

Beyond the identification sciences are certain other subfields of forensic science that assert knowledge that was neither borrowed from extant basic science nor rigorously tested by forensic experts before being offered to courts. These fields include the areas of fire, arson and explosives, gunshot residue, comparative bullet lead analysis, and aspects of forensic pathology. These areas of claimed expertise share with the forensic identification sciences the irrational reliance on unspecified, unsystematic “experience” coupled with plausible-sounding arguments as the nearly exclusive bases for their hypotheses.

The nonscience forensic sciences have failed on several levels. They are scientific failures in the sense that science (either in substance or in methodology) played little more than a

rhetorical part in the development of these fields (Saks 2007). The word and the accoutrements of science were exploited to sell these fields to courts and to the public (Cole 2001), but they did not trouble themselves to build upon a foundation of substantive scientific knowledge or to employ the methods of science to self-critically evaluate their wares. They are also technological failures. The specific techniques that have been developed and deployed in these fields make unknown numbers of errors—both false negatives (failing to detect what they sought to detect, or to identify what they sought to identify) and false positives (seeming to detect something that did not exist, mistakenly identifying an innocent person or object as having been connected to the crime under investigation). Yet despite these core failures, the nonscience forensic sciences have been extraordinarily “successful” as measured by their admission into trial courts and their apparent acceptance by the public. The reasons for acclaim in the public domain is a subject beyond the scope of this review, although prominence on nightly television programs and in feature-length movies must constitute part of any explanation. Acceptance among courts and lawyers, however, is somewhat easier to document, though possibly as difficult to explain.

In this review, we focus on the nonscience forensic sciences and consider the quandary of how such profoundly flawed disciplines could achieve such worldly success. In examining this puzzle, we explore three fundamental concerns. The first section considers the basic scientific failure of these disciplines. This consideration includes a brief survey of their historical origins and a detailed analysis of their compositions, in an effort to identify how they came to be and to understand how they persist. The second section considers the law’s role in perpetuating these disciplines. Although courts provide the primary markets for the experts coming from these fields, they have done virtually nothing to evaluate critically the bases for the opinions that representatives from these fields offer into evidence. Finally, the third section looks forward and asks how these nonscientific forensic sci-

ences might be made scientific and what such a world might look like.

SCIENTIFIC FAILURES

Although science is not monolithic, the many forms science assumes share certain characteristics. Perhaps foremost, good science has a strong element of formality associated with it. This formality permits what might be the key attribute of the scientific method: reproducibility. Although postmodern sensibilities are wont to stress objectivity as a keystone of science, virtually all science strives toward what Sir Karl Popper (1949) called intersubjective testability. In a nutshell, this means that the essential characteristics of any scientific test can be understood well enough to be replicated by another group of researchers. Researchers must specify all the details employed in testing the hypothesis of interest so that outsiders can know just what was done. This requires clear statements of operational definitions, full exposition of research methods employed and statistics used, and strict limits on the degree of subjectiveness relied upon in reaching judgments or conclusions. The scientific method, therefore, is not merely a way of testing hypotheses; it permits outsiders to know how the tests were conducted and allows other researchers to repeat and confirm them.

From the law’s perspective, as a consumer of scientific opinion, knowing what methods were used to support various theories or hypotheses should be an imperative. Without knowing in detail how (and whether) a field tested its hypotheses, it is impossible for courts to know how valuable, if valuable at all, expert opinions are that come from that field. Although anecdotal experience sometimes produces valuable knowledge, without verifiable tests it is impossible to know when it does so.

Ostensibly scientific expert opinion, therefore, must be evaluated on the basis of fundamental principles of good hypothesis testing. Unfortunately, as this section describes, much of forensic science fails this basic test. The nonscience forensic sciences have largely

developed outside of mainstream science and are not premised on the basic sciences. Moreover, their essential hypotheses have been subjected to few if any serious tests, and indeed, their primary hypothesis of individualization may be inherently impossible to test. Absent real testing that can be replicated by other researchers and independently verified by courts, forensic identification science is not really science at all.

No Basic-Science Origins

Although forensic science is routinely defined as “the application of science to problems of law,” this description does not fit the subfields of forensic science that are concerned with individualization. Although knowledge from organic chemistry can be brought to bear in identifying what drugs, poisons, or medications might be discovered in a corpse found at a crime scene, what knowledge from conventional sciences like biology or chemistry or physics support the notion of individualization? To what university science department does one go, what course does one take, to learn the basic scientific research that undergirds the notion that no two firearms impart indistinguishably similar markings or that reveals how one knows when a questioned writing was made by one writer rather than another?

Even a cursory review of major university course catalogs quickly reveals that nothing could supply a knowledge base for the forensic identification sciences. Only a relative handful of schools nationally have forensic science departments, and most of those are devoted to teaching the technologies of the past rather than testing the limits of these technologies or developing their scientific foundations. More telling, and more troubling, forensic issues are rarely taken up by academic scientists of any kind. Statisticians have generally ignored the subject, although they regularly investigate parallel problems, such as signal-detection theory (Swets 1973, Phillips et al. 2001).

In the nineteenth century, science was as likely to be found outside the university struc-

ture as it was to be found inside it. This was particularly true in the United States, where the university itself was not fully formed until the twentieth century (Rudolph 1990 [1962], Thelin 2004). It was not unusual for serious science to be explored by wealthy amateurs. Benjamin Franklin and Charles Darwin are particularly well-known exemplars of this model. But, over time, science became increasingly institutionally based, and by the mid-twentieth century most science was done by professionals associated with universities. These departments housed communities of scientists, set standards of acceptable practice, and educated succeeding generations of researchers. Perhaps most importantly, these departments embodied and promoted a culture of scientific curiosity and adventure. In the best science departments, this combined a commitment to well-demonstrated methodological and statistical practices and encouraged innovation and creativity in hypothesis formation. Although science departments achieved such cultural ideals imperfectly and not invariably, the incredible success of the academic model in the twentieth century is undeniable.

Most of the forensic identification sciences, however, missed the school bus. They never joined the university system. Although hardly surprising, but worse still, they instead became an integrated component of the police apparatus. They became an instrument of law enforcement, largely controlled by police technicians and their superiors. Innovation, such as it existed, was done by nonscientifically trained, albeit mostly well-meaning, bureaucrats. As novel technologies came into being, a bureaucratic mindset controlled their application. They were not subjected to systematic empirical testing before being offered as testimony—or, as Cole (2007) puts it, they followed a testify-first-validate-later approach (see also Cooley 2004, Pyrek 2007). Whereas the scientific ideal promotes constant questioning, the bureaucratic inclination is the reverse. Forensic identification science embraced the bureaucratic ideal of perpetuating the status quo (Wilson 1989). Substantial questioning,

testing, rethinking, and innovation in handwriting or toolmark identification, for instance, would not only mean that those trained in the old ways would be in danger of becoming obsolete; it also would mean that much of what they had testified to in the past, on which courts had relied, might come under a cloud of doubt.

This unfortunate state of affairs was compounded by the realities of judicial practice and reasoning, in two basic respects. First, the legal culture itself is inconsistent with the scientific perspective. In law, *stare decisis* operates as the core working premise, a view in fundamental tension with the idea of scientific progress. The habit of judges to defer to prior decisions disinclines appellate courts to revisit possible (or actual) errors by trial courts in any given case, and it leads trial judges to admit species of evidence that appellate courts had approved in the past, regardless of how flawed that type of evidence can be shown to be with current knowledge. Second, lawyers had, and continue to have, little training in the basic methods of science. They do not know how to think about validation of science claims or how to frame an assessment of such claims. In particular, most lawyers have little or no training in statistics, a key element of the applied sciences of which forensics should have been a part. Many lawyers simply could not distinguish between real science and pretensions to science. Lawyers—on both sides—who did (and do) know the difference often were too overwhelmed by resource limitations and caseload pressures to be able to mount a meaningful challenge to the evidence. Moreover, scientifically naive courts were sometimes romanced by the claim that they were being presented with science by scientists—terms appropriated by police forensic personnel even though much of what they do would be described in more conventional scientific workplaces as the work of technicians. At the same time, other courts were complicit in what they realized was a parody of science but that they, too, saw as useful in achieving convictions.

DNA profiling is the exception that proves the rule. DNA was discovered in 1953 by scientists working in academic science departments

(Morange 2000). The value of this discovery for law enforcement was obvious. If everyone has virtually unique DNA, and a copy of every person's DNA is contained in every cell of his or her body, then a technology that can profile DNA found in material left at crime scenes should be a boon to law enforcement. Once DNA technologies permitted individual identifications, both scientists and lawyers immediately appreciated its power. In just over a decade, DNA technology progressed through several iterations and various advances, including RFLP (restriction fragment length polymorphism) techniques, PCR (polymerase chain reaction) testing, and mitochondrial DNA analysis. Today, although academic scientists continue to monitor and refine DNA forensics, the technology is largely mechanical and operated by technicians who have little understanding of the scientific or statistical bases underlying the protocols they follow. The history of forensic DNA is almost a paradigm of how basic science can contribute systematically and systemically to forensics.

If forensic individualization science had emerged from normal science, its approach and its techniques probably would resemble DNA typing, with its measurement of attributes, sampling of variation in populations, and statistical bases. Error rates, probability levels, confidence intervals, and so on would be natural parts of what developed. The elements of subjectivity in forensic examination would themselves be topics of research, to understand both their operation and how to tame them.

The nonscience forensic sciences, however, followed a separate path. They were not incorporated into university science departments and ever since have remained orphaned. They became instruments of police and prosecutors. In this capacity, they have always been evaluated by their success in court. By this measure, they have been a sensation. But their success as science, measured by such pedestrian concepts as reliability and validity, is considerably less certain. It is uncertain because it is unknown. Their basic hypotheses have not been tested in any rigorous or systematic way. By any reasonable

measure of science, then, these forensic fields are failures.

Impossibility of Individualization

Possibly the most fundamental error made by the identification branches of the nonscience forensic sciences is the statement of the basic hypothesis of their fields. They repeatedly claim the ability to individualize. In other words, a basic operating premise of their expertise, they claim, is the ability to say that some trace evidence (e.g., a latent fingerprint, markings on a bullet, or a shoe imprint) was made by a particular source (e.g., finger, gun, or shoe) to the exclusion of all other possible sources in the world. This is a form of Sir Karl Popper's white swan problem (Saks 1998), which he used to demonstrate the limits of the inductive method (i.e., expertise based on simple experience or observation). Popper (1959) asked his readers to consider the hypothesis that "all swans are white." This hypothesis might be tested by the experience of observing 1000 swans, all of which are white. But no matter how many swans are observed, the possibility always remains of finding a nonwhite swan. Upon finding, say, a black swan, a researcher would have to modify his hypothesis, perhaps positing that "99.9% of swans are white," a hypothesis similarly susceptible to rigorous test but immune to certain proof. Popper concluded that even though hypotheses could not be conclusively proven true, they could be rigorously tested—that is, scientists could attempt to falsify them. Most working scientists would recognize this Popperian model as representing the essence of what they do on a day-to-day basis (as a community, though not necessarily as individuals). The assertion of the ability to individualize not only was not tested; the very framing of the hypothesis is highly vulnerable to falsification. When falsification inevitably occurs in each nonscience forensic science subfield—by "different" being mistaken as the "same" or the "same" being mistaken as "different"—rather than revise the core hypothesis to a more plausible and probabilistic one, practitioners either

completely ignore the evidence of falsification or seek ways to explain away the falsifications in order to keep insisting there are none.

In fact, it appears that "[i]ndividualization is unique to forensic science" (DeForest et al. 1983, p. 7). Conventional science has nothing to say about individualization. Individualization was invented by police investigators to meet a criminal justice system need, namely, to try to discover who had committed a crime (Thornton 1975) and to secure a conviction. In seeking to establish pinpoint linkages between crime scene evidence and known exemplars, the forensic identification sciences sought to do what no other field had ever tried to do and about which no other field has developed any basic science knowledge. Although individualization is the centerpiece of numerous forensic science subfields, a recent review argues that no theoretical or empirical basis for individualization exists, and none is likely to come into being in the foreseeable future (Saks & Koehler 2008). Interestingly, numerous forensic scientists recognize that a scientific basis for such identification does not exist and that practitioners are making a leap of faith when they offer individualization testimony in court (Inman & Rudin 2001; Stoney 2001, 1991; Bowers 2008; and others).

Forensic individualization begins with the recognition that the fine physical features of objects and the marks they leave (bitemarks or toolmarks, for example) vary. But rather than measuring that variability and carefully assessing the degree to which it reduces the pool of objects that could have been the source of the questioned evidence, systematically and rigorously taking into account the quality and amount of the questioned markings, all but one or two of the forensic identification sciences (those one or two being DNA typing and microscopic hair comparison) quickly adopted a belief in uniqueness: that no two objects could leave indistinguishably similar residua or markings. Much rides on this belief. If true, it eliminates the need to collect systematic data to assess variability. It leads to the belief that if two marks are indistinguishable they must have originated

from the same object “to the exclusion of all others in the world.” And, by a more obscure route, it has led many forensic examiners to believe that the possibility of error is (or borders on) zero.

The notion of uniqueness originated in the nineteenth century with Quételet, who hypothesized that “nature never repeats,” offering the product rule of probability theory to suggest great odds against such repetition (Cole 2001). Bertillon, a police records clerk in Paris, applied the idea to the classification of prisoners and invented anthropometry. Though never proven empirically and impossible to prove statistically, several generations of forensic scientists have repeated the mantra. The founders of each forensic identification subfield invoked the product rule (see Saks & Koehler 2008). Thus, forensic identification science has taken a fundamental insight (variability produced by random effects) and exaggerated it into an unfounded but strongly held faith in uniqueness and (virtual) freedom from error. For example, Cummins & Midlo (1943 [1961]) undertook an extensive effort to prove that no two fingerprints could be indistinguishably alike, but in the end conceded failure: “[I]t is impossible to offer decisive proof that no two fingerprints bear identical patterns.” Speaking of the futility of relying on probability theory, they wrote, “It is unfortunate that this approach carries the implication that a complete correspondence of two patterns might occur...” More recently, Stoney (1991) constructed an article around the question, “What made us ever think we could individualize using statistics?” In the field of handwriting identification, Harris (1958) showed that many people with the same name sign their names indistinguishably similarly, a vitally important disconfirmation of a core assumption that has been almost completely ignored by forensic document examiners.

Let us be clear: Efforts to identify the person who is the source of a latent fingerprint deposited at a crime scene will likely be successful much of the time, most of the time, or nearly all of the time. That it is impossible to

say how often is symptomatic of the larger problem. No one knows how often errors occur or what factors promote or prevent errors. That is because systematic empirical research on the foundations of handwriting identification, on its performance in practice, or on what should constitute best practices had been all but nonexistent for a century. Research showing foundational claims to be untrue or unprovable was simply ignored by the field (see, e.g., Cole 2005, Cummins & Midlo 1943 [1961]). Research suggesting improved practices was disregarded in favor of traditional practices (Risinger et al. 2002). When the first *Daubert* challenge was posed to asserted fingerprint expertise, proponents found themselves with so little research or any other real evidence to offer in support of their claims that in an effort to close the knowledge gap they launched two studies in mid-pretrial of that first challenge. This is a remarkable situation for a field to find itself in after a century of admission to courts.

The absence of knowledge has been filled with assertions supported by little more than intuition, anecdote, and *ipse dixit*. In the place of systematic research, the field developed a set of faith-based beliefs that are more akin to a religion or an ideology than to a science but that have largely succeeded in making all possible difficulties disappear, at least in the eyes of courts, fact finders, and the public. These unproved (and sometimes unprovable) assertions include such notions as that the error rate of fingerprint identification is zero; that the theory and the technique are flawless; that errors that do come to light always result from misapplication of the technique, not from the technique; that no two people have fingerprints that are indistinguishably alike (or confusable one for another); and so on.

These beliefs sidestep the real issues that examiners and courts need to be concerned about. The claim of uniqueness—be it true or false, provable or unprovable—is largely irrelevant. By analogy, assuming all faces are unique does not mean that one is never mistaken for another. The real problems are that latent prints are typically partial (a fraction of a fingerprint) and/or

of varying quality (smudged, distorted, overlaid with other marks). They are compared by examiners who work with no objective guidelines and who must use their judgment to reach a conclusion that is required to be of absolute subjective certainty (if a conclusion is reached). Like DNA typing, the problem facing fingerprint examiners is actually to estimate the population frequency of the attributes observed in the latent print and the known print and to estimate how many equally good matches exist in the population. The likelihood of mistakes (almost certainly) depends on numerous factors that have not been systematically studied and that are dealt with haphazardly in practice (Budowle 2006). Research to answer the questions implied in this paragraph would bring us closer to knowing how well examiners perform the task for which they are testifying or, as *Daubert* and *Kumho Tire* conceive the inquiry, the task at hand in the case before the court. For now, courts and the public will hear only the exaggerated, faith-based answers.

What we have just said about fingerprints can be multiplied across all the conventional forensic identification sciences. The good news is that research has begun, in fits and starts, more in some fields than in others. Some fields might be found to be so weak that they must be limited to no more than an investigatory role, not courtroom testimony. (This seems to have been the fate, for example, of voiceprints and compositional bullet lead analysis.) Others might be found to be more solid than they are currently regarded by informed commentators (e.g., Thornton & Peterson 2008). Some day better answers will be known. But not yet.

Consider the so-called ACE-V technique, claimed primarily by fingerprint examiners, though starting to be picked up by other identification sciences. The acronym stands for Analysis, Comparison, Evaluation, and Verification. The analysis step has the examiner look at the quality and quantity of detail that is present, often aided by magnification of some sort. The comparison step usually calls upon the examiner to compare the trace evidence to a candidate source. At the evaluation step, the exam-

iner forms an opinion regarding individualization, though, as noted above, this is not guided by any formal standards or statistical criteria. Finally, the last stage involves the reexamination by a second examiner of the trace and reputed source material to verify the match, though that person almost always knows the first examiner's conclusion (Stoney 2008). Just about everything that is wrong with today's forensics can be found embedded somewhere within the ACE-V system. Each step calls for a subjective impression guided by no objective standards.

Latent fingerprint identification is a problem that should be readily amenable to testing, considering the amount of time and attention given to it, not to mention the availability of digital databases of fingerprints. Latent prints typically are partial fingerprints found at the scene of the crime (or other place of interest) and can vary greatly in general quality as well as in the quantity of detail available for analysis. For the analysis step of the ACE-V method, however, there has yet to be any systematic attempt to determine the frequency of the fingerprint minutiae (e.g., ridge characteristics) that are analyzed and subsequently used for comparison. Suppose an examiner finds five minutiae that "match" between the latent print and a known print. The meaning of that degree of similarity depends entirely on the frequency of occurrence in the population for each of the five characteristics and, further, whether they are statistically independent of one another. Without more, the probability of randomly finding five characteristics that match is somewhere between 0.0001 and 0.9999. Nothing in the ACE-V method permits fingerprint examiners to say what the likelihood of finding a match is. Indeed, fingerprint examiners refuse to estimate probabilities; they merely insist that their subjective judgments amount to certainties.

This discussion is not meant to suggest that the kinds of information criminalists bring to the crime-solving process are useless. They can be of enormous help in narrowing the pool of suspect persons or objects. But everyday examiners have ignored their fields' own cautionary

literature, and the claims of the fields have been woefully underresearched and oversold to courts and to the public.

Absence of Empirical Testing

Even though most of the nonscience forensic sciences had no basic science to build upon, as self-proclaimed sciences they might have conducted their own empirical research to test the beliefs on which their work and their conclusions relied. But they did virtually no such research.

As one example, after reviewing the literature in their own field of forensic odontology in search of the basis for the field's beliefs that it can individualize bitemarks, Pretty & Sweet (2001) concluded, "Despite the continued acceptance of bitemark evidence in European, Oceanic and North American Courts, the fundamental scientific basis for bitemark analysis has never been established." Risinger et al. (1989) scoured the handwriting identification literature and found essentially no research testing whether such asserted experts could do what they purported to be able to do. In the area of firearms identification, some firearms and toolmark examiners have been working to place their judgments on more systematic empirical footing (Biasotti et al. 2008). This program, while admirable, remains some distance from completion and has not been embraced by most firearms examiners. For a critical review of the field, see Schwartz (2005).

There are no systematic, rigorous, empirical research bases on which the great majority of beliefs in the nonscience forensic sciences are built. (For detailed discussion of various forensic sciences, see relevant chapters of Faigman et al. 2008.) If called upon to prove their claims, they have little or no data to marshal in their field's defense. Instead, they generally point to a guild of mutually self-reassuring examiners who have come to believe in the truth of their claims, often sounding more like a faith-based religion than a data-based science (see, e.g., Moenssens 1998).

As Giannelli & Imwinkelried (2000) have noted, in each area of the nonscience forensic sciences, "little rigorous, systematic research has been done to validate the discipline's basic premises and techniques, and in each area there is no evident reason why such research would be infeasible." For example, until fairly recently, the field of fire and arson investigation relied on intuition and imagination to develop a set of what came to be accepted as indicators of arson. They came to believe that crazed glass, spalled concrete, certain types of burn patterns, and so on indicated fires that had been deliberately set. Untold numbers of cases, both criminal and civil, were decided on the strength of such experts applying those beliefs and opining on whether they thought a particular fire was arson or accident. Eventually, those beliefs were put to empirical tests in which buildings were set afire in ways that simulated either arson fires or accidental fires. By comparing the effects of the arson versus accidental fires on windows, walls, burn patterns, and so on, these conceptually simple experiments revealed that many of the accepted indicators of arson did not, in fact, distinguish arson from accidental fires. In light of the research findings, the field began to abandon those incorrect beliefs and adopt more accurate beliefs based on systematic empirical testing (Lentini 2008). Of course, some investigators learn of or come to accept the new discoveries more slowly than others. And an unaddressed concern is all of the past verdicts of criminal and civil liability or denial of insurance claims that were made in error, relying on the expert opinions that relied on untested beliefs that now are recognized to be false.

Little Use of Scientific Methods

Scientists have a way of using the scientific method in much of what they do. For example, in trying to determine what testing procedures produce the most accurate results, they might try out different ones and compare the results obtained. The ones that consistently facilitate more accurate results would be adopted as the standard procedure, the state of the art.

Similarly, systematic experiments can be conducted on different training procedures to determine which work best and which enable trainees to acquire the most effective skills most efficiently. Those are then adopted. These sorts of efforts are largely absent from the nonscience forensic sciences, and in their place are committed beliefs in the validity of the unproven.

Most immediately serious and troubling, whereas conventional scientists design studies and procedures to maximize the contribution of the phenomenon under scrutiny and minimize the contribution of expectations and biases (in particular, using blind or masked tests so as not to be misled into seeing what they want or hope or expect to see), forensic scientists—alone among those who regard themselves as scientists—do not employ such procedures and insist that they are not affected by such context effects (Risinger et al. 2002). Most scientists would suspect that knowing other inculpatory evidence in a case could distort the conclusions of someone judging and interpreting other evidence in a case, but forensic scientists have insisted that they can will themselves not to be vulnerable to this phenomenon.

However, recent research by Dror and colleagues no longer permits such denials. For instance, Dror et al. (2006) found that four out of five fingerprint experts who had several years earlier identified two prints as a match later reached different conclusions on their second encounter with those same prints, when each examiner was told that others had declared the prints to be from different persons. Although context effects are usually found to be small but persistent (usually leading to error rates of a few percent at most), 80% of Dror et al.'s fingerprint examiners changed their opinions when presented with misleading contextual information. (See also Dror & Charlton 2006.) A recent meta-analysis of these data found an effect size of $r = 0.40$, which is quite large for context effects (Dror & Rosenthal 2008). A famous fingerprint error by the FBI, in which several of their most senior examiners one after another erroneously identified a latent print in a terrorist bombing in Madrid as belonging to an

Oregon lawyer who had no connection at all to the bombing, has been attributed partly to context effects (Stacey 2004). (For descriptions of other cases of fingerprint identification errors, see Cole 2005.) Such research results and incidents suggest that criminalists should employ the same kind of blind examination procedures that are used widely in other fields. To date, however, forensic scientists continue with business as usual.

Error Management

Normal scientists do not deny error. Quite the contrary: They have been obsessed with the measurement and reporting of error at least since the Greek astronomers. Today, most scientists have made statistics—the study of uncertainty—an integral part of how they think and what they do, and they work closely with statisticians. Research studies report probability levels, specify confidence intervals, present error bars on graphs, and so on. They not only measure and report error, scientists study error, explore its sources, and work to manage it. In striking contrast, most forensic scientists, especially criminalists, pretend that error does not exist in their work. This is most plain in the work of fingerprint examiners, who insist that their error rate is zero and who declare the reporting of a qualified identification (meaning an identification accompanied by a probability statement of less than 100% certainty) to be an unethical act.

The forensic expert's crabbed understanding of error is well illustrated in the case of *U.S. v. Allen* (2002), in which the issue was the admissibility of footwear identification techniques. The court observed as follows:

With respect to the rate of error in the ACE-V [i.e., analysis, comparison, evaluation, and verification] process, [the expert] testified that the error rate of the process itself is zero, meaning that based upon the science, the shoe either did or did not make the impression. Any error that does occur, according to [the expert] is caused by examiner error in the application of

the process or by examiner error in reaching a particular conclusion (p. 862).

A theoretical error rate of zero should be reassuring to no one. No technology has a zero error rate or failure rate in practice, whether the fault lies with the method, the operator, or most likely both. To say that the suspect shoe either did or did not make the impression is a metaphysical statement that addresses neither the real concerns about the ability to associate any given shoe with such a mark nor the various risks of error associated with any inference related to whether it did or did not. Courts, after all, operate in the real world, not the world of metaphysical imaginings. To even suggest that footwear identification techniques have a zero error rate, based on some theoretical belief that every shoeprint is unique, is simply rhetoric, not science.

Claims of zero error rates are particularly ironic given that all individualization sciences (with the exception of DNA typing) rely exclusively on each examiner's subjective judgment to assess the commonness or rarity of features (of fingerprints, toolmarks, etc.), because no efforts have been made to determine the population frequency of those features (except of DNA). Indeed, DNA evidence, the gold standard in forensic identification, is itself susceptible to an assortment of potential sources of error. These include, among others, errors in collecting biological material, mislabeling of evidence, accidental contamination of samples, and, sometimes, intentional fraud. All the potential sources of error that plague DNA identification also threaten other kinds of forensic evidence. But with the latter, we have little or no idea of the random match probabilities. As Stoney (2008) has pointed out, "The criteria for absolute identification in fingerprint work are subjective and ill-defined. They are the product of probabilistic intuitions widely shared among fingerprint examiners, not of scientific research" (p. 360). It may seem paradoxical that a field whose work is so lacking in objective measurement, and whose decisions turn on subjective judgment, would eschew the

analysis of probability and error. But for several reasons it should not be surprising. When one has no objective measurements, the calculation of error and probabilities is more difficult. Moreover, this scientifically odd behavior likely is a product of the adversary system. Forensic scientists are trapped in the crossfire of prosecutors who want evidence to be as airtight as it can be and defense attorneys who will look for any daylight that might be seeping through the government's case. It has been suggested that over the decades forensic examiners have been pressured to make statements as extreme as they can get—100% certainty, zero error rates, identification to the exclusion of all others in the world—not because such statements grow out of their science, but because they serve the needs of those who use their work (Saks 1998).

Prominent forensic science thinkers have questioned these implausible conventions. Thornton & Peterson (2008) have observed: "[T]hough individualization is clearly the goal toward which forensic science strives, it can be achieved only in a probabilistic sense, of reducing uncertainty to the smallest possible amount . . ." (p. 7). "Behind every opinion rendered by a forensic scientist there is a statistical basis. We may not know what that basis is, and we may have no feasible means of developing an understanding of that basis, but it is futile to deny that one exists" (p. 22). Champod & Evett (2001) have argued that "there is, at present, a major contradiction between the scientific status that is claimed and the operational paradigm to which its practitioners subscribe. This contradiction is exemplified by a recent statement by . . . SWGFAST [Scientific Working Group on Friction Ridge Analysis, Study and Technology]: 'Friction ridge identifications are absolute conclusions. Probable, possible, or likely identification are outside the acceptable limits of the science of friction ridge identification'" (p. 432). Champod & Evett pose these challenging questions: "Is a statement of an 'absolute conclusion' compatible with scientific reasoning? Is the denial of probabilistic reasoning compatible with a scientific pursuit?"

Moreover, claims of no error and no possibility of error fly in the face of studies revealing errors in casework (Cole 2005), in proficiency testing (see Thornton & Peterson 2008), and in other studies (e.g., Dror et al. 2006, Dror & Charlton 2006, Kam et al. 1997). (For more extensive reviews, see Volumes 4 and 5 of Faigman et al. 2008.) The irony may be that by denying the possibility of error, forensic scientists have denied themselves the opportunity to understand better the sources of error and to manage error better. Lest one think that errors hurt only defendants, bear in mind that by setting a decision threshold so high that one declares an identification only when one is “100% sure,” the incidence of false negative errors is increased (Phillips et al. 2001, Haber & Haber 2003).

Surprisingly, forensic science expert evidence appears to be one of the leading causes of erroneous convictions, second only to eyewitness testimony (see **Figure 1**).

As discussed in regard to the ACE-V method above, forensic individualization science involves two fundamental steps. The first step is to compare a questioned item of evidence to an exemplar from a known source and to judge whether they appear so alike that they can be said to match. The second step is to assess the meaning of that reported match: What is the probability that the questioned and the known originated from one and the same source? Different risks of error are present at each step. The risk of error in the first step is that a reported match between a questioned and known sample might not really match. Even if the method used to compare questioned and known samples were flawless, an error could occur if, for example, one of the samples had been mislabeled or mixed up with a different sample. The risk of error associated with the second step is that the reported match may be accurate but arose through coincidence (the random match problem) rather than because the samples share a common source. The risks of error at both steps affect the ultimate inferences that can be drawn about the identification evidence in a case. (For

more thorough discussion, see Thompson & Cole 2006.)

Both risks are significantly underresearched. As to the first step, existing standards and procedures do not provide sufficient protection from erroneous conclusions that two marks are indistinguishably alike—that is, that they match when in fact they differ. Few, if any, criminalistics subfields have objective standards for deciding whether two patterns match. That determination is left to the subjective judgment of each examiner. For example, consider Stoney's (2001) discussion of fingerprint examination standards:

How much correspondence between two fingerprints is sufficient to conclude that they [are the same pattern]...? An adequate answer... is not currently available. The best answer at present... is that this is up to the individual expert fingerprint examiner to determine, based on that examiner's training, skill, and experience. Thus, we have an ill-defined, flexible, and explicitly subjective criterion for establishing fingerprint identification... Any unbiased, intelligent assessment of fingerprint identification practices today reveals that there are, in reality, no standards (p. 329).

The lack of objective standards helps explain some disturbing findings from the small body of research on pattern matching by forensic scientists that has been conducted to date. In some tests, examiners disagreed with one another about whether various images matched (see proficiency studies summarized in Faigman et al. 2008). In other tests, examiners who agreed that two patterns matched disagreed (sometimes dramatically) on what constituted the match (Evelt & Williams 1996). Examiners differ not only in their ability to perceive pattern similarity and differences, but also in their thresholds for calling matches.

As to certainty of forensic identifications, FBI forensic scientist Bunch (2000) has noted: “[I]t must be observed that there is no rational or scientific ground for making claims of absolute certainty in any of the traditional

identification sciences, which include fingerprint, document, firearms, toolmark, and shoe and tire-tread analysis.”

ILLUSORY JUDICIAL OVERSIGHT

The courts, and the legal profession more generally, have been remarkably ineffective in policing forensic science. With only a few notable exceptions, courts have not closely evaluated the scientific bases for forensic identification and have not been responsible for exposing the flaws in any of the forensic science fields that have, over time, been abandoned owing to their lack of validity. For reasons that will become clear in this section, the failures of the courts are intimately connected to the failures of the forensic sciences.

The reasons for the courts’ empirical ineffectuality are likely manifold. Three systemic factors, in particular, appear to explain much regarding the courts’ failures in this area. The first, and most basic, is the cultural divide that separates law and science. Outside of legal content areas such as intellectual property, patents, and the environment, lawyers have little training in the basic precepts of research methodology and statistics. Although law school could be described as a glorified liberal arts education, it generally does not include courses in research methods and statistics. Even where such courses are available, they are taken by a small percentage of students. The second factor inevitably affecting the law’s insouciance is the standard by which courts judge the probative value of expert evidence. During most of the twentieth century the courts used admissibility tests that were either explicitly deferential to the professional fields that supplied the expertise or operated deferentially as a practical matter. Although modern admissibility standards appear to mandate a less indolent review of proffered expertise, little has changed on the criminal side of courts’ dockets. This may be a function either of courts’ inclination to grandfather in expertise traditionally admitted or of some predisposition that favors evidence offered by prosecutors. The third factor is the ironic situation

that because nonscience forensic science is not scientific, there are few defense experts available to testify. Most scientific fields are highly dynamic and often hotbeds of debate and disagreement about the meaning of their research and the content and reach of their theories. Real science teaches doubt. Good faith disagreement among scientists spills over into the courtroom and provides fodder for the adversarial process. Although the courtroom is an imperfect forum for resolving scientific disagreement, this dialectic at least produces considerable information for courts to ponder. Nonscience forensic science creates no similar content in large part because group consensus substitutes for science.

These systemic challenges might combine to help make judicial ignorance of the methods of science seem easy to understand. Yet it still remains quite puzzling. Although law is a distinct institution with its own goals and objectives, it constantly interacts with the world and institutions around it. The law receives input from a variety of sources, digests it through the legal process, and applies the output with the expectation of effecting some result. These steps require judges to have extraordinarily broad understanding of an assortment of professional disciplines. For instance, judges routinely rely on historical sources, particularly in constitutional cases, and it would be remarkable if they either eschewed responsibility for understanding it critically or employed it in plainly ignorant ways. Yet when it comes to science, and particularly statistics, judges pause and sputter, wondering whether it is truly part of their responsibility to know the details of scientific methods. On occasions when judges are required to think in a critically scientific way, the results can be inane.

A particularly trenchant example of this inanity can be found in the case of *U.S. v. Havvard* (2000). The *Havvard* court attempted to justify its decision to admit fingerprint expert testimony in a written opinion that inadvertently does just the opposite. The court employed the admissibility standard set forth in the Federal Rules of Evidence and famously

explicated in the case of *Daubert v. Merrill Dow Pharmaceuticals, Inc.* (1993). *Daubert*, as discussed further below, establishes a scientific threshold for the admission of ostensibly scientifically based expert evidence. In particular, the *Daubert* court instructed trial courts to consider, among other things, whether the basis for scientific expert opinion had been adequately tested and whether the testing had been methodologically sound (with help in that evaluation coming from peer review and publication), had acceptable error rates, and was generally accepted in the pertinent field. Remarkably, the *Havvard* court took these indicators of good science and converted them into crime investigation and litigation markers. On the issues of falsification and whether the techniques are testable or have been tested, the court wrote that “[t]hey have been tested in adversarial proceedings with the highest possible stakes,” as if the stakes and not the research design, methods, and analysis are the key to sound empirical research. On the issue of “peer review and publication,” the court confused the process of fundamental knowledge building in a science with the practice of examiners checking each other’s everyday casework. The court stated that “the methods of identification are subject to peer review. [A]nother qualified examiner can compare the objective information upon which the opinion is based and may render a different opinion if warranted. In fact, peer review is the standard operating procedure among latent print examiners.” Although there was no research to review, there still was peer review, the court suggested. The court excused the absence of published research by saying again that “latent fingerprint identification has been subject to adversarial testing for roughly 100 years, again in cases with the highest stakes possible.” On the issue of error rates, the court was told that fingerprint examiners never make errors. Although it disbelieved the government’s “breathtaking” claim that “the error rate for the method is zero,” the court made what it considered a “reasonable concession” that occasional mistakes “can of course” occur. It admitted the evidence.

Thus, in *U.S. v. Havvard* (2000), the court substituted courtroom debate for scientific research in an effort to justify admission. Following this strange exercise, the judge declared this expert evidence to be “the very archetype of reliable expert testimony under [*Daubert*]” (p. 855).

In another fingerprint case, *U.S. v. Llera-Plaza II* (2002), the court acknowledged the gaping absence of empirical research testing the claims of fingerprint examiners, but it searched for some reason, any reason, to admit the expertise. The judge in *U.S. v. Cline* (2002) took the view that the Supreme Court could not have wanted the nonscience forensic sciences, such as asserted fingerprint identification, to be excluded, and therefore, notwithstanding *Daubert*’s apparent commands, soundness (validity, reliability) of the evidence could not be the touchstone of admission. Consequently, in the *Cline* court’s view, given the serious shortcomings in the scientific foundations of fingerprint expert evidence, alternative, less rigorous, criteria would have to be employed in order to facilitate admission. Put differently: If the student cannot pass the test, the test has to be made as easy as it needs to be to ensure that the student can pass it. A review of the many cases in which courts wrestled with challenges to fingerprint expert testimony summed up those courts’ opinions (in which they could not find bases for admission that comported with the requirements of *Daubert* and *Kumho Tire*, but nevertheless found various ways to rule the testimony admissible) as amounting to “a catalog of evasions” (Faigman et al. 2008, chapter 32, § 32:3).

These cases, and the many others that might be cited (see, generally, Faigman et al. 2008), well illustrate courts’ limited understanding of the basics of science. These limitations, however, were little on display until fairly recently. As noted above, in 1993 the Supreme Court changed the fundamental orientation federal judges use for evaluating the admissibility of expert evidence. Prior to *Daubert*, courts that used any articulated test at all used one that was deferential to the professionals from the pertinent field, a test most famously associated with

the case of *Frye v. U.S.* (1923). The *Frye* test queried whether the knowledge or technique on which the scientific opinion depended was “generally accepted in the particular field” from which it came. (For a detailed review of *Frye* and its comparison with *Daubert*, see Saks & Faigman 2005.) *Frye* and similar tests seemed to require no special scientific acumen. *Daubert*, in contrast, changed the primary focus from what was accepted by those who practiced in the field to whether the methodology and principles underlying the expert opinion were reliable and valid. The *Daubert* analysis, which was subsequently codified in amendments to the Federal Rules of Evidence in 2000, requires a fair degree of scientific sophistication to be conducted well. Courts lacking such sophistication, however, have turned the test into something of a caricature.

When *Daubert* was first decided (see Saks & Faigman 2005), no consensus existed among courts or commentators about whether the new test loosened or tightened the judicial filtering of such testimony. It is probably fair to say that, initially, most observers believed the test was less stringent. But as courts applied the test, it turned out to be more stringent because many expert opinions that had been routinely admitted turned out, upon inspection, to have little data and poor research methods behind them. As applied, therefore, *Daubert* seemed to raise the threshold for scientific evidence. Courts, however, made an exception for the nonscience forensic sciences, for which the test has usually been applied so as to make it as undemanding as necessary to facilitate admission.

This strange mosaic can be understood only by considering two major planes, each of which is itself somewhat paradoxical: what *Daubert* does (rather than what it says) and what judges do when they apply that law (something that legal realists conceived as the law on the books versus the law in action). (For a more thorough review and analysis of these issues, see Saks & Faigman 2005; Faigman et al. 2008, chapter 1.) On the plane of law, *Frye* required that a proffered expertise be generally accepted within its field, whereas *Daubert* requires that

the basis for proffered expertise be demonstrably sound, valid, and reliable. Thus, *Daubert* ought more readily to admit expert testimony that has a sound basis even if it has not yet gained general acceptance, whereas *Frye* would more readily admit expert testimony that had come to be generally accepted even though it has no demonstrably sound basis. No family of asserted expertises falls more clearly into the category of simultaneously being generally accepted (at least within their own orbit) and yet lacking any demonstrably sound basis than the nonscience forensic sciences. Thus, logically, these fields ought to have been more readily admitted by *Frye* and more readily excluded by *Daubert*. On the plane of judicial behavior, however, although the switch from *Frye* to *Daubert* led to more intense scrutiny and more frequent exclusion of expert testimony in civil cases, something quite different occurred on the criminal side.

Judges have, in general, bent over backwards to evade the application of *Daubert* when conscientious application would lead to the exclusion of any of the nonscience forensic sciences. For example, upon realizing that asserted handwriting identification expertise (FDE, or forensic document examination) had no sound scientific basis, the court in *U.S. v. Starzecpyzel* (1995) declared:

Were the Court to apply *Daubert* to the proffered FDE testimony, it would have to be excluded. This conclusion derives from a straightforward analysis of the suggested *Daubert* factors—testability and known error rate, peer review and publication, and general acceptance—in light of the evidence adduced at the *Daubert* hearing (p. 1036).

Seeing where *Daubert* led, the court looked for a way to get off of *Daubert*'s path. In the end, the court reasoned that FDE lacked scientific validity; because it lacked scientific validity, it was not subject to *Daubert* review; and because it was not subject to *Daubert*, it was admissible. Were it a sound science, it would have been admissible; because it was not a sound science,

it was admissible. This solution was followed by numerous federal courts until it was implicitly overruled by the Supreme Court's decision in *Kumho Tire v. Carmichael* (1999), holding that all fields of asserted expertise, whether labeled science or anything else, had to meet serious and appropriate criteria of soundness or be excluded.

One would have expected that after nearly a century of admission as evidence in courts, when finally called upon to defend the scientific foundations of their claims in court, fingerprint examiners would be able to roll wheelbarrows filled with studies and data into court to prove the bases for their claims. But they had nothing. Their expert testimony had been presented in American courts since early in the twentieth century, having been admitted first by courts that applied no real test of validity, and then by courts that relied on those earlier courts as precedent of a sort (Saks 1998). After the Supreme Court's decisions in *Daubert* (1993) and *Kumho Tire* (1999), challenges began to be laid (see Saks & Faigman 2005). Amazingly, the first empirical studies subjecting the basic claims of fingerprint examiners to empirical testing did not occur until the midst of the pretrial phase of the first of those post-*Daubert* challenges, *U.S. v. Mitchell* (2000). The judge in that case simply ruled, without a written opinion, that fingerprints were admissible. The first court to use the record developed for *Mitchell* in a formal opinion ruled fingerprint expert opinions on individual identity to be inadmissible (*U.S. v. Llera-Plaza I* 2002). The same court reversed itself several months later, admitting the testimony, though it continued to find that virtually no research existed supporting the claims of fingerprint experts (*U.S. v. Llera-Plaza II* 2002).

State courts have been similarly wide-eyed in their assessments of the nonscience forensic sciences. In a case involving a challenge to the admission of expert testimony on microscopic hair comparison, for example, the Kentucky Supreme Court, in *Commonwealth v. Johnson* (1999), purported to apply a *Daubert* analysis to admissibility. Because the court had no

relevant data of any kind in the record before it, the proffered expertise could not meet any of the *Daubert* criteria—except perhaps general acceptance. But the court also could find no evidence of general acceptance, and so it relied on its own past cases from decades earlier for the conclusion that microscopic hair identification examination passes general acceptance muster. It turns out, however, that those earlier cases said nothing about general acceptance. The Kentucky Supreme Court acknowledged this fact but reasoned that those courts “must have” implicitly found general acceptance or they would not have admitted the hair identification evidence. The flaw in this reasoning was that Kentucky had not adopted *Frye* or any other version of general acceptance until after those cases had been decided. Through this sleight of hand—forming grounds for admission out of thin air—the Kentucky Supreme Court held that microscopic hair identification satisfied *Daubert*.

In another state case, involving asserted expertise in voiceprint identification, *Alaska v. Coon* (1999)—the very case by which the Alaska Supreme Court adopted *Daubert* as its test of admission—the court remanded the case to the trial court to take additional evidence in anticipation of adopting and applying *Daubert*, but no court involved in the case at any level built its holding on any relevant research studies or other data. Undaunted, in another magical judicial opinion, the Alaska court conjured the bases for admission out of nothing.

The problem has not merely been judges who cannot bring themselves to follow the law to its logical conclusion if doing so will result in an outcome that makes them nervous. Judges in criminal cases are not likely to have the opportunity to hear many challenges because lawyers in criminal cases are much less likely than lawyers in civil cases to mount challenges to the other side's scientific expert evidence (Risinger 2000). This is a reflection of the familiar sociological and institutional differences between civil and criminal cases.

Some courts have accepted the command of federal or state versions of *Daubert* and ruled as

required by those tests. This was true especially in regard to asserted handwriting identification expertise. Following *Kumbo Tire*, most federal courts that thoughtfully examined this evidence followed the path of *U.S. v. Hines* (1999), which held that experts could not testify regarding the identity of the author of a questioned document. These courts limited the expert to “show-and-tell” descriptive statements regarding similarities and differences between the known and unknown exemplars. Indeed, in some circumstances, courts excluded the testimony of forensic document examiners entirely (*U.S. v. Saelee* 2001, *U.S. v. Fujii* 2000; for a more complete discussion, see Faigman et al. 2008, chapter 33). Several courts ruled that fingerprint identification expert opinions on identity are mostly inadmissible (*U.S. v. Llera Plaza I* 2002) (though later reversed) or completely excluded (*Maryland v. Rose* 2007).

But at the end of the day, these cases are small islands of inadmissibility surrounded by a sea of admission. As one commentator summed up this seascape: “It seems that the only standard the courts are requiring of forensic science is that it be incriminating to the defendant” (Frisbie & Garrett, 1998, p. 56, quoting John Marshall Law School Professor Melvin B. Lewis). In essence, the nonscience forensic sciences are granted exemption from serious scrutiny. Despite the hopes of the framers of our Constitution, the decisions of individual judges are not made in a vacuum insulated from pressures of various kinds. This is not to say that the pressures are improper or illegal. The spotlight focused by the news media on an unusual (even if unusually correct) decision, or the shrillness of pleas from the government when it loses, or considerations internal to the judicial branch can impel judges to decide in ways other than what they believe to be correct. For a noteworthy example, consider the opinion of the judge in *U.S. v. Green* (2005), a case involving firearms identification, where the proffered expert insisted he could identify the suspect weapon “to the exclusion of all others in the world” despite being unable to offer any basis for believing that

to be true. In admitting the testimony, albeit only partially, the court wrote:

I reluctantly [admit the proffered expert testimony] because of my confidence that any other decision will be rejected by appellate courts, in light of precedents across the country, regardless of the findings I have made. While I recognize that the *Daubert-Kumbo* standard does not require the illusory perfection of a television show (CSI, this wasn’t), when liberty hangs in the balance—and, in the case of the defendants facing the death penalty, life itself—the standards should be higher than were met in this case, and than have been imposed across the country. The more courts admit this type of toolmark evidence without requiring documentation, proficiency testing, or evidence of reliability, the more sloppy practices will endure; we should require more (p. 109).

Indeed, we should.

WHAT IS TO BE DONE

In Hans Christian Andersen’s famous tale, “the Emperor walked in the procession under the lovely canopy, while all the crowds in the street and all the people at their windows said, ‘Heavens, How marvelous the Emperor’s new clothes look!’” The crowd, in fact, saw that the emperor had no clothes, but feared saying so: “No one wanted it thought that he could not see anything, as that would make him somebody who was either very stupid or badly fitted for his position.” A little child, however, spoke up: “But he has nothing on!” The crowd realized that the child was right and all began to shout: “He has nothing on!” The emperor realized it, too. He “cringed inside himself, for it seemed to him that they were right, but he thought like this: ‘I shall have to go through with the procession’” (Andersen 1959, pp. 30–32).

Many nonscience forensic scientists are in the same situation as the emperor. They have “nothing on.” At this point in time, of course, they continue in the procession as testifying

experts because, like the emperor, they cannot admit the reality. To stop would reveal them to be “either very stupid or badly fitted for [their] position.” Change, if it is to occur, must come from outside the nonscience forensic sciences.

There are two courses of action that might lead to wholesale reform in the practice of forensic science: First, the courts need to become better consumers of expert testimony and demand better methods than so far have been forthcoming; second, mainstream academic scientists must join the effort to test forensic hypotheses. Although many other reforms might accompany these two, both are essential to any true changes. Otherwise, these naked experts will continue to march on, with their “chamberlains . . . behind [them] carrying the train that was not there at all.”

Gatekeeping Courts

In some respects, the problems endemic in the forensic sciences must be understood in terms of market failure (Faigman 1999). The manufacturers of forensic knowledge produce inferior products, and the buyers (largely courts) continue to purchase them. Reform, therefore, can be imagined in regard to both the supply and the demand of the forensic product. Although it probably does not matter greatly which factor is revolutionized first, true change cannot take place unless the practices of both sellers and buyers change. We begin with the buyers.

To a large extent, the revolution has already occurred in the law, but many of the revolutionaries still do not realize how pervasive the reforms have been, at least in principle. *Daubert* largely put in place the procedural framework that would permit courts to see nonscience forensic science for what it is. The key paradigm shift wrought by *Daubert* was to require courts to consider the methods and principles that supported expert evidence, not simply inquire about its general acceptance among some insular group of true believers (Faigman 2000, Saks & Koehler 2005). The gatekeeping metaphor was first described in *Daubert* itself.

Largely explicit in the decision, the Court contemplated that judges were charged with the responsibility to understand science well enough to measure the scientific validity (i.e., “evidentiary reliability”) of proffered expert opinion. This responsibility led Chief Justice Rehnquist to complain that *Daubert* would require federal judges “to become amateur scientists” (*Daubert*, p. 601). And indeed they must (Faigman 2006).

Science is such a pervasive force in modern society that judges are essentially failing their obligations if they persist in their scientific ignorance. This task, in fact, is not profoundly difficult, at least in the forensics area. In most cases, recognizing the weaknesses of nonscience forensic science is not rocket science. Much of the expert opinion propounded in court has not been tested beyond mere anecdotal experience and is not supported by adequate research designs or basic statistical testing. Moreover, texts and treatises are available that detail the numerous and profound failings of these nonsciences (see Faigman et al. 2008, Giannelli & Imwinkelried 2007).

Just why the courts have so utterly failed to carry out their gatekeeping duties is difficult to say. We can only speculate about possible reasons for this dereliction. The two main possibilities are lack of knowledge or lack of will. We believe the latter is more likely, for several reasons. First, as noted, the failures of the nonscience forensic sciences are so profound and so elementary that they are just not that difficult to understand. Second, the same judges who demur from their gatekeeping obligations when prosecutors proffer bad forensics are vigorous gatekeepers when considering bad expertise proffered by plaintiffs in civil cases. It is hard to believe that judges who minutely criticize plaintiffs’ epidemiological proof cannot understand that prosecutors’ latent fingerprint evidence lacks scientific validation. Finally, so many of the judicial opinions dealing with forensic evidence employ such tortured logic that it is impossible to believe that courts are blind to the empirical realities of this evidence. It appears, therefore, that lack of will is a better explanation than plain ignorance.

There are, of course, many credible reasons why courts should lack courage in this context. Foremost, perhaps, they fear that excluding any forensic expertise that lacks a scientific basis will result in wholesale exclusion of substantial portions of prosecutors' cases. This could lead to the disruption of large numbers of cases and lead prosecutors not to pursue other cases. Moreover, it would potentially throw into doubt hundreds, if not thousands, of past cases, possibly setting off a firestorm of habeas petitions seeking to overturn convictions based on now-excluded forensic evidence. In individual cases, where courts are convinced from other evidence that a defendant is guilty, they will see no need to scrutinize the forensic evidence closely because (if it too inculpates) they will assume it is correct. Evaluating science, for its own sake, is not the purpose of judicial proceedings. Finally, courts legitimately see no present alternative to this proof. Whatever its weaknesses, courts probably think it has to be admitted until something better comes along.

From the perspective of courts, therefore, the current state of the art of forensic science cannot be jettisoned until a new state of the art is available. Yet, it also appears that without courts raising serious doubts about current practices, there is little pressure for the forensics field to change course. As a consequence, the status quo will persist until something intervenes. If neither the courts nor the forensic experts are likely to change things, the only group left is the mainstream scientific community. It is there that real reform in forensics can take place.

Scientific Gate-Makers

In light of the prevailing practice in forensic science, there is little reason to believe that the field will soon begin a comprehensive and rigorous research program. Indeed, even if the field generally recognized the need for such a program, it is not at all clear that it has the wherewithal to even begin this work. Most professional forensic experts are employed by state or private laboratories that apply forensic technology, not produce it. Existing laboratories are

not likely to have the time, resources, or inclination to vigorously test the technology they employ daily. Such an effort would certainly appear to them as a waste of time and money. Why test what already works, at least insofar as "what works" is measured by its acceptance in court? More problematic still, today's forensic scientists are generally not trained as researchers; they are technicians. A study of crime laboratory personnel found that fewer than 1% had PhD-level research degrees and that only about 3% had earned a master's degree (Furton et al. 1999). The overwhelming majority of forensic professionals simply do not have the statistical or methodological training to do original research. Hence, if good science is to be brought to bear on forensic hypotheses, mainstream academic scientists will have to take the laboring oar.

At least two ingredients are necessary to interest academic scientists in the study of forensic issues: money and relevance. The first is relatively easily explained, if not readily obtained. Research, quite simply, follows the money. Historically, little government financing has been available for basic forensic science research. Agencies such as the National Science Foundation and the National Institutes of Health do not include forensics as part of their mission statements. Probably the single agency most devoted to funding forensic research is the National Institute of Justice (NIJ), an arm of the Justice Department. Thus far, NIJ has not been successful in funding cutting-edge research. Finally, because academic scientists have not considered forensics to be within their bailiwick, they have not asked for money to do this kind of research. Clearly, before academic scientists can be expected to reform forensic science, they must be convinced that the subject is relevant and important to their day jobs.

Although the question of the relevance of forensics to various mainstream scientific communities is too large to canvass fully here, just brief contemplation on the subject provides a clear answer. Forensics provides a treasure trove of research possibilities. Statisticians, in particular, would find a surplus of hypotheses to test in

areas such as latent fingerprints, firearms identification, and handwriting analysis. Indeed, conventional statistical subjects such as pattern-recognition and signal detection theory could be tested and applied to forensic problems. Moreover, many scientists would find forensics to be a fertile area to test general theories from their disciplines. The work of Dror and colleagues—psychologists interested in studying cognitive biases—is a particularly good illustration of this. The list matching forensic subjects to established academic departments is limited only by one's scientific imagination.

In the short term, mainstream science has the potential to reform forensic science through the efforts of individual scientists (e.g., Zabell 2005) and the contributions of select scientific organizations. In particular, the National Research Council of the National Academies of Science has played, and could continue to play, a pivotal role in both evaluating the state of the art of forensic science and in making recommendations for the future (e.g., National Research Council 1992, 1996). In the long term, however, only with substantial funding and the concerted efforts of dedicated individuals will the nonscience forensic sciences become scientific. Perhaps one day forensic science will become another subject in the vast scientific curriculum of major research universities.

CONCLUSION

We have explored the state of nonscience forensic science—the absence of basic-science origins, unsupported assumptions, exaggerated claims, lack of empirical testing, little use of scientific method, pretermission of error. And we have discussed the interplay of nonscience forensic science with illusory judicial oversight, as well as the indifference of normal science toward forensic science.

Let us conclude by imagining what forensic science would look like if it had grown up in the world of conventional basic and applied science, as typically found in the departments and laboratories of universities and industry and many government agencies. And, there-

fore, what might it look like if in the future it could be set on a path toward that world, whether by its own initiative, under the incentives of courts conscientiously applying *Daubert* and its progeny, by the restructuring of the way forensic services are organized (Koppl 2005), or by normal scientists becoming seriously interested in the subject?

The underlying assumptions of forensic identification would be subjected to intense questioning and empirical testing. In the course of this work, researchers would report what they tested, how they tested it, and what they found. Their results—whether good, bad, or indifferent—would be reported with equal candor. Because many of these areas (handwriting, firearms, tiremarks, etc.) had no basic science from which to draw, these researchers would be building the science that would be the foundation of forensic science.

The individualization techniques that emerged would probably resemble DNA typing, with its measurement of attributes, sampling of variation in populations, and statistical bases. Error rates, probability levels, confidence intervals, and so on would be natural parts of what developed. The elements of subjectivity in forensic examination would themselves be topics of research, the better to understand their operation and how to tame them. These scientists, like most, would recognize problems of experimenter and observer effects; their own research would use procedures designed to prevent such effects, as would the procedures they designed for use in crime labs.

Those doing this research would be doctoral scientists and their graduate students, assisted by novices, some of whom would become the scientists and technicians who would later staff the crime labs applying the techniques being developed. Incompetence, error, and fraud would be minimized by the various cultural and organizational features of the scientific enterprise that have evolved, though sometimes such problems would still arise. Deliberate fraud is a career-ending act. At the end of the day, errors (whatever their cause) are revealed by the failure

of findings to replicate. All this work would provide the science for forensic science.

As the knowledge became ready for “prime time” (and as newer knowledge and developments improved upon older), it would be adopted for casework in crime labs. It is anyone’s guess how crime labs would be organized in such a world, but they likely would inherit some of the culture of the conventional scientific world along with the knowledge base drawn from that world. This might include an ethos of candor, a dedication to data above all else, and therefore a hostility toward incompetence or fudging or fraud. It might also include closer and more reciprocal relationships with basic researchers, attentiveness to

new findings, and an alertness to problems in need of further basic and applied research. Because much of the knowledge probably will have been grown in universities, there would be more college courses in these subjects (indeed, there would be a larger body of scientific knowledge to teach, more than techniques and assumptions). Personnel would arrive at crime labs with more and better education in relevant subjects. A more highly trained doctoral- and master’s-level cadre of lab directors and managers would come into being and would more readily move between the worlds of research and casework. These labs and their personnel would be both more welcoming of oversight, review, accreditation, and certification and less in need of it.

DISCLOSURE STATEMENT

Some might wonder whether the fact that the authors occasionally testify in hearings where the admissibility of some particular forensic science is at issue creates a bias. In fact, however, their research and publications have led to requests that they testify about that scholarship, just as it has led to invitations to speak to audiences of academics, judges, forensic scientists, and governmental and quasi-governmental bodies, as well as invitations to contribute papers to legal, social science, and forensic science publications.

LITERATURE CITED

- Alaska v. Coon*, 974 P.2d 386 (Alaska 1999)
- Andersen HC. 1959. *The Emperor’s New Clothes*, transl. E. Blegvad. New York: Harcourt
- Biasotti A, Murdock J, Moran B. 2008. Firearms and toolmark identification. See Faigman et al. 2008, 4:571–629
- Bowers CM. 2008. Identification from bitemarks. See Faigman et al. 2008, 4:649–743
- Budowle B, Buscaglia J, Schwartz Perlman R. 2006. Review of the scientific basis for friction ridge comparisons as a means of identification: committee findings and recommendations. *Forensic Sci. Comm.* 8(1). http://www.fbi.gov/hq/lab/fsc/backissu/jan2006/research/2006_01_research02.htm
- Bunch SG. 2000. Consecutive matching striation criteria: a general critique. *J. Forensic Sci.* 45:955–63
- Chamod C, Evett IW. 2001. A probabilistic approach to fingerprint evidence. *J. Forensic Identif.* 51:101–22
- Cole SA. 2001. *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard Univ. Press
- Cole SA. 2005. More than zero: accounting for error in latent fingerprint identification. *J. Crim. Law Criminol.* 95:985–1078
- Cole SA. 2007. Forensic science: it’s the testimony stupid. *Natl. Law J.* 30(Dec. 10):22
- Cooley C. 2004. Reforming the forensic science community to avert the ultimate injustice. *Stanford Law Policy Rev.* 15:399–446
- Cummins H, Midlo C. 1943 (1961). *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics*. New York: Dover
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)
- DeForest PR, Gaensslen RE, Lee HC. 1983. *Forensic Science: An Introduction to Criminalistics*. New York: McGraw-Hill

- Dror IE, Charlton D. 2006. Why experts make errors. *J. Forensic Identif.* 56:600–15
- Dror IE, Charlton D, Péron AE. 2006. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci. Int.* 156:74–78
- Dror IE, Rosenthal R. 2008. Meta-analytically quantifying the reliability and biasability of forensic experts. *J. Forensic Sci.* In press
- Evelt IW, Williams RL. 1996. A review of the sixteen points fingerprint standard in England and Wales. *J. Forensic Identif.* 46:49–73
- Faigman DL. 1999. *Legal Alchemy: The Use and Misuse of Science in the Law*. New York: W.H. Freeman
- Faigman DL. 2000. The law's scientific revolution: reflections and ruminations on the law's use of experts in year seven of the revolution. *Washington Lee Law Rev.* 57:661–84
- Faigman DL. 2006. Judges as “amateur scientists.” *Boston Univ. Law Rev.* 86:120–25
- Faigman DL, Kaye DH, Saks MJ, Sanders J. 2008. *Modern Scientific Evidence: The Law and Science of Expert Testimony*, Vols. 1–5. St. Paul, MN: West/Thomson
- Frisbie T, Garrett G. 1998. *Victims of Justice*. New York: Avon
- Frye v. U.S.*, 293 F. 1013 (D.C. Cir. 1923)
- Furton KG, Hsu YL, Cole MD. 1999. What educational background is required by crime laboratory directors? *J. Forensic Sci.* 44:128–32
- Giannelli P, Imwinkelried E. 2000. Scientific evidence: the fallout from the Supreme Court's decision in *Kumho Tire. Crim. Just.* 14:12–40
- Giannelli PC, Imwinkelried EJ. 2007. *Scientific Evidence*. Newark, NJ: LexisNexis. 4th ed.
- Haber L, Haber RN. 2003. Error rates for human fingerprint examiners. In *Automatic Fingerprint Recognition Systems*, ed. NK Ratha, R Bolle, pp. 339–60. New York: Springer-Verlag
- Harris J. 1958. How much do people write alike: a study of signatures. *J. Crim. Law Criminol.* 48:647–51
- Inman K, Rudin N. 2001. *Principles and Practice of Criminalistics*. Boca Raton, FL: CRC Press
- Johnson v. Commonwealth*, 12 S.W. 3d 258 (Kentucky 1999)
- Kam M, Fielding G, Conn R. 1997. Writer identification by professional document examiners. *J. Forensic Sci.* 42:778–86
- Koppl R. 2005. How to improve forensic science. *Eur. J. Law Econ.* 20:255–86
- Kumho Tire Ltd. v. Carmichael*, 526 U.S. 137 (1999)
- Lentini J. 2008. Fires, arsons, and explosions. See Faigman et al. 2008, 5:57–159
- Maryland v. Rose*, No. K06-0545 (2007). <http://www.barlib.org/BryanRoseFingerprintMemDecision2.pdf>
- Moenssens A. 1998. Handwriting identification evidence in the post-*Daubert* world. *U.K. Law Rev.* 66:251–343
- Morange M. 1998. *A History of Molecular Biology*, transl. Matthew Cobb. Cambridge, MA: Harvard Univ. Press
- National Institute of Justice. 2007. *Focus Group on Scientific and Forensic Evidence in the Courtroom*. Washington, DC: NIJ
- National Research Council. 1992. *DNA Technology in Forensic Science*. Washington, DC: NAS
- National Research Council. 1996. *The Evaluation of Forensic DNA Evidence*. Washington, DC: NAS
- Osterburg JW. 1969. The evaluation of physical evidence in criminalistics: subjective or objective process? *J. Crim. Law Criminol.* 60:97–101
- Phillips V, Saks M, Peterson JL. 2001. The application of signal detection theory and decision-making in forensic science. *J. Forensic Sci.* 46:294–308
- Popper K. 1959. *The Logic of Scientific Discovery*. New York: Basic Books
- Pretty I, Sweet D. 2001. The scientific basis for human bitemark analyses—a critical review. *Sci. Justice* 41:85–92
- Pyrek KM. 2007. *Forensic Science Under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Death Investigation System*. Amsterdam/Boston: Academic
- Risinger DM. 2000. Navigating expert reliability: Are criminal standards of certainty being left on the dock? *Albany Law Rev.* 64:99–152
- Risinger DM, Denbeaux MP, Saks MJ. 1989. Exorcism of ignorance as a proxy for rational knowledge: the case of handwriting identification “expertise.” *Univ. Pa. Law Rev.* 137:731–92
- Risinger DM, Saks MJ, Thompson WC, Rosenthal R. 2002. The *Daubert/Kumho* implications of observer effects in forensic science: hidden problems of expectation and suggestion. *Calif. Law Rev.* 90:1–56

- Rudolph F. 1990 (1962). *The American College & University: A History*. Athens: Univ. Georgia Press/New York: Knopf
- Saks MJ. 1998. Merlin and Solomon: lessons from the law's formative encounters with forensic identification science. *Hastings Law J.* 49:1069–141
- Saks MJ. 2007. Remediating forensic science. *Jurimetrics J.* 48:119–24
- Saks MJ, Faigman DL. 2005. Evidence after *Daubert*. *Annu. Rev. Law Soc. Sci.* 1:105–30
- Saks MJ, Koehler JJ. 2005. The coming paradigm shift in forensic identification science. *Science* 309:892–95
- Saks MJ, Koehler JJ. 2008. The individualization fallacy in forensic science. *Vanderbilt Law Rev.* 61:199–219
- Schwartz A. 2005. Toolmark and firearm identification. In *Psychological and Scientific Evidence in Criminal Trials*, ed. J Campbell Moriarty, pp. 12–49. Deerfield, IL: West Group
- Stacey RB. 2004. Report on the erroneous fingerprint individualization in the Madrid train bombing case. *J. Forensic Identif.* 54:706–18
- State v. Quintana*, 103 P.3d 168 (Utah Ct. App. 2004)
- Stoney DA. 1991. What made us ever think we could individualize using statistics? *J. Forensic Sci. Soc.* 31:197–99
- Stoney DA. 2001. Measurement of fingerprint individuality. In *Advances in Fingerprint Technology*, ed. HC Lee, RE Gaensslen, pp. 329–30. Boca Raton, FL: CRC Press. 2nd ed.
- Stoney DA. 2008. Fingerprint identification. See Faigman et al. 2008, 4:275–377
- Swets JA. 1973. The relative operating characteristic in psychology. *Science* 182:900–9
- Theelin JR. 2004. *A History of American Higher Education*. Baltimore, MD: Johns Hopkins Univ. Press
- Thompson WC, Cole SA. 2006. Psychological aspects of forensic identification evidence. In *Expert Psychological Testimony for the Courts*, ed. M Costanzo, Krauss DA, Pezdek K, pp. 31–68. Mahwah, NJ: Erlbaum
- Thornton J. 1975. Criminalistics: past, present, future. *Lex et Scientia* 11:1–44
- Thornton J, Peterson J. 2008. The general assumptions and rationale of forensic identification. See Faigman et al. 2008, 4:1–81
- U.S. v. Allen*, 207 F. Supp. 2d 856 (2002)
- U.S. v. Cline*, 188 F. Supp. 2d 1287 (D. Kan. 2002), aff'd, 349 F.3d 1276 (10th Cir. 2003)
- U.S. v. Fujii*, 152 F. Supp. 2d 939 (N.D. Ill. 2000)
- U.S. v. Green*, 405 F. Supp. 2d 104 (D. Mass. 2005)
- U.S. v. Harvard*, 117 F. Supp. 2d 848, 849 (S.D. Ind. 2000), aff'd, 260 F.3d 597 (7th Cir. 2001)
- U.S. v. Hines*, 55 F. Supp. 2d 62 (D. Mass. 1999)
- U.S. v. Llera-Plaza I*, 2002 WL 27305 (2002)
- U.S. v. Llera Plaza II*, 188 F. Supp. 2d 549, 565 (E.D. Pa. 2002)
- U.S. v. Mitchell*, *Crim. No. 96-407-1* (E.D. Pa., judgment entered February 2000)
- U.S. v. Starzecpyzel*, 880 F. Supp. 1027 (S.D.N.Y. 1995)
- U.S. v. Saelee*, 162 F. Supp. 2d 1097 (D. Alaska 2001)
- Wilson JQ. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books
- Zabell SL. 2005. Fingerprint evidence. *J. Law Policy* 13:143–79



Figure 1

Postmortem analysis of causes of wrongful conviction in cases of DNA exonerations.



Contents

Home Away from Home: Collaborative Research Networks and Interdisciplinary Socio-Legal Scholarship <i>Stuart A. Scheingold</i>	1
Conditionality: Forms, Function, and History <i>Sarah L. Babb and Bruce G. Carruthers</i>	13
Organizations, Regulation, and Economic Behavior: Regulatory Dynamics and Forms from the Nineteenth to the Twenty-First Century <i>Marc Schneiberg and Tim Bartley</i>	31
The Political Economy of American Indian Gaming <i>Stephen Cornell</i>	63
After Inclusion <i>Devon Carbado, Catherine Fisk, and Mitu Gulati</i>	83
Divergent Paths: Conflicting Conceptions of Employment Discrimination in Law and the Social Sciences <i>Robert L. Nelson, Ellen C. Berrey, and Laura Beth Nielsen</i>	103
Providing Expert Knowledge in an Adversarial Context: Social Cognitive Science in Employment Discrimination Cases <i>Susan T. Fiske and Eugene Borgida</i>	123
Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It <i>Michael J. Saks and David L. Faigman</i>	149
Convicting the Innocent <i>Samuel R. Gross</i>	173
The Psychology of Confessions <i>Saul M. Kassin</i>	193
Forecasting Methods in Crime and Justice <i>Richard Berk</i>	219

Undercover Policing and the Shifting Terms of Scholarly Debate: The United States and Europe in Counterpoint <i>Jacqueline E. Ross</i>	239
Jury Systems Around the World <i>Valerie P. Hans</i>	275
Women in the Legal Profession <i>Fiona Kay and Elizabeth Gorman</i>	299
The Reform of Legal Education in East Asia <i>Setsuo Miyazawa, Kay-Wah Chan, and Ilhyung Lee</i>	333
The Countermajoritarian Difficulty: From Courts to Congress to Constitutional Order <i>Mark A. Graber</i>	361
Toward a New Sociology of Rights: A Genealogy of “Buried Bodies” of Citizenship and Human Rights <i>Margaret R. Somers and Christopher N.J. Roberts</i>	385

Indexes

Cumulative Index of Contributing Authors, Volumes 1–4	427
Cumulative Index of Chapter Titles, Volumes 1–4	429

Errata

An online log of corrections to *Annual Review of Law and Social Science* articles may be found at <http://lawsocsci.annualreviews.org>