Duquesne University School of Law

From the SelectedWorks of Jane Campbell Moriarty

October 2005

Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping

Contact Author

Start Your Own SelectedWorks

Notify Me of New Work

Available at: http://works.bepress.com/jane_moriarty/4

Forensic Science

Grand Goals, Tragic Flaws, and Judicial Gatekeeping

By Jane Campbell Moriarty
and Michael J. Saks

Volume 44 • Number 5 • Fall 2005 • American Bar Association • Judges' Journal "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

Electronic copy of this paper is available at: http://ssrn.com/abstract=961441

In the last decade, a number of scientists have published articles and testified in court, explaining the ways in which they believe that some of the forensic sciences do not meet reliability standards and that laboratories make errors. The explosion of exonerations resulting from DNA technology has raised questions about the accuracy of many forensic sciences and the quality of some laboratory testing. A substantial number of these defendants can point to erroneous forensic science as a contributing cause of their wrongful convictions. In the courts, increasingly, the parties have substantial and serious disagreements about the quality of forensic science.

The adoption of the reliability standard for expert evidence in federal courts and many state courts has created a daunting task for trial judges who must grapple with any number of complex, scientific, and technical forms of evidence. Judge Kozinski first recognized the enormity of the effort in the remanded *Daubert* decision, summarizing the new mandate and remarking, "We take a deep breath and proceed with this heady task."[1] Because the Supreme Court, many state courts, and the Federal Rules of Evidence have extended proof of reliability to all forms of expert testimony,[2] courts must grapple with questions concerning forensic expert evidence that is a part of many criminal prosecutions. Make no mistake—this is a challenging task, and many judges wrestle with the proper application of reliability standards to this type of evidence. This article provides some assistance to the judges dealing with the complexities of forensic science expert testimony.

We offer three kinds of guidance concerning the proper role of forensic science. The first part of the article provides an exposition of the types and uses of forensic science frequently proffered. The second part addresses the notable questions, problems, and concerns about forensic science. The article closes with suggestions for approaches that may assist courts when confronting these questions and concerns.

Nature and Goals of the Forensic Sciences

Although three areas of forensic science are generally recognized—identification, individualization, and reconstruction—we address only the identification and individualization spheres, and focus primarily on the problems associated with individualization.

Identification/quantitation versus individualization. An important distinction exists among the forensic sciences. The identification/quantitation specialties typically are derived from conventional basic sciences, from which these forensic specialties have inherited well-developed and well-tested principles applicable to answering disputed factual questions in litigation. The parent field most often is chemistry, though principles are drawn from physics or biology. The goal of these specialties typically is to identify a substance (that is, to determine what something is) and to quantitate it (that is, to measure how much of that something there is).[3] For example, what chemical is a certain powder, and how much of it is present? Does a cadaver contain poison, what kind, and how much? These fields, based on conventional science, are considered highly reliable and are rarely challenged in court unless a new technique is presented or evidence of negligence or fraud exists.[4]

The individualization[5] specialties have quite different goals from the identification sciences. The individualization specialties seek to associate an item of evidence found at a crime scene with its unique source, to the exclusion of all others. That is not a goal shared with any conventional sciences, and thus the methods and principles of these specialties have not been derived from conventional sciences. Rather, the methods and principles employed by the individualization specialties—involving the comparison of such things as bite marks, bullets, fingerprints, footwear, hair, handwriting, and so on—were invented by and for these specialties themselves.

The identification/quantitation specialties thus differ from the individualization specialties in several important ways: In goals and origins and in a panoply of additional theoretical, empirical, and technological areas. Though the forensic sciences often define themselves as the application of science to problems of law, this description is more obviously true of identification specialties than of individualization specialties.

DNA typing is the exception. It is the first individualization science derived from traditional science. It thus serves as an illuminating counterpoint to and model for the other individualization sciences.

Areas not asserting individualization: identification and quantitation. General forensic testing includes the analysis of such liquids as blood or urine to determine, for example, blood type and enzymes. This type of testing also analyzes solids, such as clothing or other materials, to ascertain the presence of gunpowder residue or blood stains on the fabric. The following sections discuss the various methods commonly used in crime laboratories in the areas of identification.

Serology testing. Laboratories often must first determine whether a reddish-brown stain is indeed blood and, if it is, whether it is of human origin. After presumptive tests suggest that blood is present, additional confirmatory immunological tests[6] are then performed to determine if it is

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

Electronic copy of this paper is available at: http://ssrn.com/abstract=961441

human. Presumptive tests for blood tend to be very sensitive (needing small amounts of blood to generate a positive result) but are also prone to false positive results (commonly encountered materials unrelated to blood give results that cannot be distinguished from those for blood). In contrast, confirmatory tests rarely give false-positive results but require larger amounts of material for testing and more often give rise to false negative results (a tested sample really does contain human blood but the test reports that it does not). Most testing laboratories will subject samples that test positive for blood with either presumptive or confirmatory tests to genetic typing.

In addition to analyzing blood, serological tests can also be performed to determine the presence or absence of other bodily fluids, such as semen, urine, saliva, and other physiological fluids.[7] Semen and blood can be identified with virtual certainty, but fluids such as saliva and urine can be only presumptively identified.[8]

Chromatography. Gas and liquid chromatographs are common fixtures in crime laboratories throughout the country. They are used to detect the presence of various components in a sample. For example, bodily fluids can be tested by chromatography to determine whether they contain poisons, drugs, or alcohol.[9] Chromatograph results are both qualitative, by identifying a component's presence, and quantitative, by measuring the amount of that substance.[10]

Microscopy. Various types of microscopes are used in crime laboratories, including the commonly employed stereo-binocular microscope, which provides a magnified image of the object as it actually appears, allowing the viewer to see the three dimensions of length, breadth, and depth.[11]

Microscopes are used to separate small amounts of trace materials that are relevant to a crime scene (such as material in a pant cuff), or to look closely at pieces of larger items to determine whether they contain bloodstains or markings.

Although polarizing microscopes allow examiners to study physical and chemical properties of specimens, comparison microscopes are employed to permit examination of two specimens at the same time.[12] By employing an optical bridge, comparison microscopes allow various samples to be viewed in a single field, for example, to determine whether two fibers are similar or not.

Of a more sophisticated nature are scanning electron microscopes (SEMs), which produce magnification in excess of 300,000 times, far greater than other microscopes, which generally do not exceed the power of 2,000.[13] SEMs use a beam of electrons directed at the item in question. When the beam strikes the surface, the interaction generates secondary electron emissions, which create a micrograph of the image that appears in shades of gray, black, and white, and permits observations of objects and characteristics that were not visible even with the strongest optical microscope.[14]

Areas asserting individualization. In principle, all forensic individualization specialties assume uniqueness in the markings and objects of interest: fingerprints, handwriting, hair, bite marks, tool marks, etc. These unique differences are assumed to be the product of random variations in the creation of, or changes in, the objects. These variations are explained, if at all, by the notion of the multiplication rule in probability theory, which suggests that where different objects vary on numerous attributes, the probability of any two of them sharing the same attributes is extremely small.[15] They assume, further, that although not all features are "individualizing" (some features are "class" characteristics, shared by numerous persons or objects in the class), at least some are unique to the particular object or person. Part of the claimed skill is to distinguish class from individual features and not to be fooled by subclass characteristics. Judgments of inclusion or exclusion require careful comparisons of features. When marks (or material) are found at a crime scene or elsewhere that might have been left by the perpetrator or crime victim, examiners compare that questioned evidence to those known to belong to a suspect or victim to determine whether the two can be associated or excluded.

Exclusion is easier to determine than association. If two marks differ in any unexplained way, they cannot share a common source. Inclusion requires a judgment that (1) two sets of features are indistinguishably alike and (2) those similarities are shared by no other person or object. None of these specialties, except DNA typing, possesses a database that permits the calculation of an objective probability of the likelihood that other persons or objects share those same features (known to statisticians as the probability of a coincidental match).

Jane Campbell Moriarty is associate professor of law at the University of Akron in Ohio, concentrating on expert evidence and legal ethics. She is the author and editor of *Psychological and Scientific Evidence in Criminal Trials* (1996, ann. supps. 1997–2005).

Michael J. Saks is professor of law and psychology and a fellow of the Center for the Study of Law, Science, and Technology at the Arizona State University. He is co-author and co-editor of *Modern Scientific Evidence* (2005) and co-editor of the *Annotated Reference Manual on Scientific Evidence* (2005).

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

Except where noted below, the basic concepts and methods employed in the earliest days of a specialty continue to be used today. Advances have taken place in peripheral technologies (e.g., improved photographic techniques or chemicals for making latent fingerprints visible), not in the core theory by which judgments of association or exclusion are made.[16]

DNA evidence—nuclear and mitochondrial. By far the most soundly based of the forensic individualization sciences is nuclear DNA (nDNA) testing; mitochondrial DNA (mtDNA) testing is less discriminating. Although DNA testing does not declare a match between a known and unknown sample, it expresses the likelihood in statistical terms that the sample in question came from someone other than a suspect.[17] Since DNA testing was first used in the late 1980s, three methods commonly are used to generate DNA profiles: the restriction fragmentation length polymorphism method (RFLP), the amplification method using polymerase chain reaction (PCR), and the short tandem repeats test (STR).[18] Almost all DNA testing laboratories currently perform STR tests due to the tests' sensitivity, discriminatory power, and ability to yield results quickly. These tests take DNA and create a numeric inventory of which genetic markers are present at several—usually thirteen—different locations on human chromosomes. Comparisons can then be made between the sample in question and a known sample to see if they have the same sets of observed markers. If the samples do not match, the known sample can be eliminated as the donor of the DNA in question. This is how DNA is able to effectively exonerate people. If the samples appear to match, then the scientist can calculate the probability that the sample in question came from a genetic donor other than the known source.

Over the last two decades, DNA has gained widespread acceptance in both scientific and legal literature as a reliable form of scientific evidence, although like any other scientific endeavor, it is not infallible.

Recently, mtDNA has entered the forensic arena. First used to identify remains, it is now used to test hair, bones, and teeth to determine identity.[19] Unlike nDNA, mtDNA is not a unique identifier, since mtDNA is shared by maternal relatives. Although siblings will typically be indistinguishable at the level of their mtDNA sequence, the test is useful where nDNA is not available.[20] Although the process for creating mtDNA profiles is very similar to that used in generating STR profiles from nDNA, mtDNA does not have the same probative value—the frequency of the most common mtDNA sequence is one in twenty-five, and the frequency of a common nDNA profile is estimated at one in ten billion.[21] To date, several courts have found mtDNA sufficiently trustworthy to be admitted.[22]

Fingerprints. The accepted trustworthiness of fingerprint comparison has led to the use of the word "fingerprint" to describe an individualization, a "DNA fingerprint," for example. Fingerprint comparison has been used in courts for a century[23] and, until recently, its ability to individuate had never been challenged; there is an almost universal belief that each fingerprint is unique and fingerprint comparison is an almost error-free science.[24] In the last few years, however, critics have challenged the claim of infallible comparisons.

A fingerprint is made when the friction ridge skin on the fingertip leaves an impression. These friction ridges are the corrugated skin on fingers, palms, and soles; they leave a "print" when a medium from the skin—such as body oils, blood, or sweat—touches a surface and transfers the friction ridge pattern. Prints that are not readily visible—latent prints—require the use of a powder or chemical that adheres to the medium to make them visible.[25] Fingerprints can also be lifted with tape from a surface and transferred to a card, a method that can also record surface details—such as scratches—in addition to the fingerprint.[26]

Fingerprints taken from subjects are generally quite different from latent fingerprints. The "ten-print identification" is an inked and rolled image of all ten fingers and has far more ridge detail than latent prints, which may include only a partial print of any one of ten fingers and may be smudged or otherwise of poor quality.[27]

Fingerprint comparison until recently had never been challenged for its ability to individuate.

A fingerprint examiner might conduct a point-to-point comparison between the latent print and the rolled print to determine whether or not to declare a match. Contrary to many crime shows, notably *CSI*, a computer does not declare a match. Rather, it provides a list of candidates ranked by a comparison algorithm; an examiner then compares the prints visually.[28]

Fingerprint examiners compare the rolled print with the latent print, looking for points of similarity, generally comparing ridge characteristics. Some examiners go further, looking at the shapes of the ridges or the location of pore ducts.[29] Examiners sometimes describe the process as "ACE-V": analyze, compare, evaluate, and verify.[30] Verification consists of having a

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

second examiner review the conclusions of the first examiner to check whether they agree. An analysis generally is not an independent, blind analysis of the prints.

The final step is to make a judgment about identification—whether the latent and known prints originated from the same person. There is a great deal of debate about the number of points of similarity one must find to declare a match, and the required number of points varies by country, ranging from eight to sixteen points.[31] Currently, neither the courts nor accepted standards used by fingerprint professionals in the United States require a specific number of points of comparison to declare a match. In contrast to other forms of scientific individualization evidence (such as DNA), the examiner does not speak in terms of probabilities but declares a positive identification, often expressed as "100 percent certainty." In fact, the International Association for Identi-fication prohibits member examiners from offering "possible, probable, or likely conclusions."[32]

Firearms and tool-mark comparison. A vast array of weapons, bullets, and knives are used in crimes. Firearm and tool-mark examiners testify to their opinions regarding whether a particular gun is the unique source of marks on bullets or other ammunition components, whether a specific tool was used to pry open a window, and (it sometimes is claimed) whether a particular knife was the weapon that produced wounds on a decedent. Firearms identification deals with the tool marks that are acquired by ammunition components such as bullets, cartridge cases, and shot shells while being fired or worked through the action of a firearm.[33]

Firearms examiners look at striated tool marks consisting of patterns of scratches, or striae, produced by the movement of an ammunition component through the weapon (for example, marks produced on a bullet when it is fired through a gun barrel). Examiners also look at impressed tool marks produced by the perpendicular action of a part of a gun on an ammunition component (for example, firing pin impressions created when the firing pin of a gun strikes the primer of a cartridge case). Although firearms examination is grounded on the assumption that each individual gun produces tool marks with unique individual characteristics, all tool marks produced by a particular type of gun (a Glock 40, for example) share the same class characteristics. The tool marks produced by a batch of guns of a particular type (a particular batch of Glock 40s, for example) might all share subclass characteristics in place of, or in addition to, individual characteristics.[34]

With nonfirearms, numerous test tool marks often must be made to approximate the angle and pressure used to produce the tool marks on an object recovered from the crime scene.[35] The examiner seeks to determine whether sufficient microscopic features are present possessing sufficient clarity and definition on which to base an opinion of identification.[36]

As with other types of forensic individualization methods, tool-mark and firearm comparisons rest on the foundational belief that each pattern of markings is unique and thus can be compared with other markings to determine whether a specific weapon was used in the crime.

Hair comparison. The two methods by which a known and unknown sample of hair can be compared are the visual method—historically widely used and relied upon in court—and the mtDNA method, which compares the mtDNA in questioned and known hair samples. Although far more cases have relied upon visual rather than mtDNA comparison, there is widespread recognition of the former's limitations[37] and the substantial potential for error.[38]

In the visual method of hair comparison, which has changed little since its origin, an examiner views sets of hairs using an optical microscope. The examiner can look at nearly two dozen different physical attributes (color, structure, diameter, weathering, etc.) of each hair. The examiner must make a judgment about whether any two *sets of* hairs are similar or different on any given attribute, similar or different in total, and, ultimately, whether the hairs come from the same person or different persons. The subjective judgment required is further complicated by the following facts: hair can come from a variety of body locations (head, arms, chest, pubis, legs); hairs from different parts of the same body vary; two (or more) hairs from the same person can appear to have come from different persons; and two hairs from different persons can appear to have come from the same person. Because of those intra-individual variations, it is necessary to have a sufficiently large and representative sample of hair[39] to make a valid judgment. Using standard microscopy techniques also can aid examiners to infer whether the hair is human, whether it is from a male or female, and what the race is of the person who is the source.[40]

Microscopic hair comparison is regarded by forensic scientists as insufficiently certain to permit an examiner to assert an opinion that two sets of hairs came from the same person (although some come very close to offering such opinions, and others *do* offer them). Examiners can report that they have found them to be "similar" or "matching in all microscopic detail."[41] Microscopic hair comparison appears to have a substantial rate of error.[42]

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

In recent years, DNA testing has supplemented microscopic comparison, and nDNA typing can be done where the root is still attached to a hair. Where only hair shafts are available, mtDNA testing can be used because mitochondria are found outside the nuclei of cells. As described above, however, mtDNA is not as discriminating as nDNA. But given the potential for error inherent in visual hair comparison methods, mtDNA testing has begun to replace microscopic-comparison testing as the preferred standard.

Handwriting identification. This is the oldest of the forensic sciences, having first been offered, and sometimes admitted, in American courts before the middle of the nineteenth century. Handwriting examiners were not widely admitted until the beginning of the twentieth century, and not without expressions of judicial skepticism until after the Lindbergh kidnapping case in 1936.[43]

Following the Supreme Court's decision in *Daubert*, the field of handwriting identification has once again been confronted with challenges and the imposition of judicial limitation and exclusion. Handwriting examiners subscribe to theories involving the acquisition and retention of fine motor skills—which one might expect to have been borrowed from such fields as psychology or physiology. These theories, however, actually are untested assumptions passed down through generations of examiners and may or may not comport with relevant scientific research.

Examiners' essential beliefs are that no two people write alike, no one person writes the same way twice, and thus no two writings are ever identical. Handwriting examiners must contend with the problem that variation exists not only among writers but within each writer. They must judge whether two writings that are not identical represent variations within a single writer's range or are the work of two different writers. By comparing questioned writing with known writing, examiners try to draw inferences about whether the writing is genuine or forged. The field states however rarely can accurately identify the writer of a forged writing[44] and try to associate "natural" (not forged or disguised) writing with its author. Handwriting examiners traditionally made remarkably strong claims of reliability and validity for their judgments, though no research was conducted to test those claims until late in the twentieth century—most of it prompted by *Daubert*.[45]

Bite mark comparisons. Forensic dentistry began its career by identifying the remains of victims of mass disasters such as fires and airplane crashes, by comparing dental records to the largely intact dentition of the victims. Such methods result in identifications about 20 to 25 percent of the time. As a group, forensic dentists until the mid-1970s were unsure of their ability to identify the source of bite marks left in the flesh of victims of violent crimes and were cautious about testifying in court. But such opinions became increasingly common—a result based more on the courts' eagerness for the testimony than on the soaring confidence of odontologists in what they offered.[46]

Identification of biters is made through a comparison of a cast of the suspect's dentition with an image or cast of the bite wound or directly with the bite mark itself. Tooth shape and positions change over time, but not usually in the brief period between a crime and its investigation. Complicating the process of accurately associating biters with bite marks are the facts that only a limited amount of dentition is recorded in a bite mark and that the flesh in which the mark was made changes shape, particularly at the time the bite is being made (as during a struggle) but afterward as well.

Though forensic dentists have performed their work for many fewer decades than the other individualization specialties, they already offer a larger body of research and more self-critical analysis than the other individualization specialties. The result of that in the short term is that wide disagreement appears to exist among practitioners; in the long term, however, the field is likely to generate a better-informed understanding of its capabilities and limitations than will other specialties.

Shoe and foot impressions. Marks left by shoes or bare feet often are found at crime scenes and may also associate suspects with or exclude them from the crime scene. Shoes can leave prints (a flat image on a hard surface) or impressions (a three-dimensional image in a soft surface). Shoes vary by class characteristics (size indications and patterns manufacturers place on their soles), by subclass characteristics, and by features that are more individualizing (patterns of wear or random damage to the sole).[47] Photographs or molds capture the images left by the shoes so they are available for later comparison with a suspect's shoes—linking the marks at the crime scene with a shoe in the suspect's possession. A different and more controversial kind of comparison, referred to as "Cinderella analysis," is an effort to link the *feet* of a perpetrator to a crime by comparing impressions left at the crime scene by a shoe's insole pattern, or by bare feet, with shoes owned by a suspect or with the suspect's foot.[48] A few footprint examiners go one step further and claim to be able to associate a shoe print or impression at a crime scene with the foot of the person who left it.

None of these methods is regarded

 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

as an established, recognized forensic science, and few examiners specialize in these comparisons.[49] Shoe-print and footprint identification provide a good illustration of some of the principles of forensic individualization. Shoe size is a class characteristic. A suspect whose shoe size is different from a shoe print found at a crime scene can be excluded from having left that print. A suspect whose shoe size is the same may be considered within the group of suspects but cannot be more narrowly linked to the crime scene (which requires more than class characteristics). A particular wear pattern might be unique, or it might be found on the soles of numerous people's shoes; therefore, it might or might not link a suspect to the crime scene. Because databases of such patterns do not exist to guide shoeprint examiners, they must rely on their intuition and experience.

Bullet lead comparison. Compositional analysis of bullet lead (CABL) is based on the idea that batches of lead used to make bullets contain unique combinations of seven trace elements: arsenic, antimony, tin, copper, bismuth, silver, and cadmium. If two bullets contain the same ratios of these elements, an expert may infer that they originated from the same source. The central premise is the same as that of other forensic individualization sciences: all units of comparison are unique, as are all batches of lead. With bullet lead, however, a single "pour" can produce a goodly number of bullets, which may end up being distributed through a vast or unknown geographic area.

Beginning in the 1950s, courts admitted expert testimony on bullet lead composition that led to inferences that a defendant committed a crime: for example, a bullet found in a victim's body might share trace element composition with a bullet found in the defendant's possession. Recently, the validity of the foundational claims for this kind of expert testimony have been scrutinized by metallurgists and statisticians, in a series of studies and reports. Most prominently is one published by the National Academy of Sciences (NAS),[50] concluding that CABL's claims are unfounded. The NAS panel concluded that the methods of measuring the amounts of trace elements were sound, but that the assumptions and inferences based on measurements and leading to the "expert opinions" expressed in court were unsupportable.

Expert testimony in this area has varied widely in experts' conclusions drawn from two bullets that are "analytically indistinguishable." Some testified that this meant the bullets "could have come from" or "came from" the same "source" or "batch." Others testified that the bullets "could have come from the same box," "could have come from the same box or a box manufactured on the same day," or "must have come from the same box or from another box that would have been made by the same company on the same day." Even the most modest of these opinions was not supported by the data.[51]

The NAS panel made a number of recommendations to the Federal Bureau of Investigation (FBI) regarding the bureau's bullet lead analyses and expert testimony based upon the tests. The suggestions included that the FBI

- Adopt single protocols for the chemical and the statistical analysis and make sure they are followed in every case.
- Publish the details of what it does so that others can fully understand its results and techniques.
- Use more appropriate statistical tests to determine whether two bullets can justifiably be said to be "analytically indistinguishable."
- In offering conclusions, avoid overbroad statements ("Witnesses should not say that two analytically indistinguishable bullets come from the same melt, production run, or box, or were made on or about the same day. None of these assertions . . . can be justified by the available data.").
- Acknowledge uncertainties ("The nature of the distribution of elements in lead batches is such that two bullets from the same batch can differ chemically and two bullets from different batches can be indistinguishable.").

In September 2005, the FBI decided that it would no longer use bullet lead comparison in solving crime cases. Although the FBI itself concluded that the manufacturing of bullets was too variable to make matching bullet lead composition reliable, the FBI laboratory director inexplicably stated, "We stand by the results of the reports we have already issued."[52]

The experience with bullet lead evidence—poor science, ready admittance by the courts, and ensuing expert testimony that created exaggerated impressions in the minds of fact finders—provides a cautionary lesson about other forensic science evidence, which is the focus of the next part of this article.

Analytical Gap between Data and Opinions

Traditional and contemporary fears about expert evidence and the law's efforts to resolve them. Although courts embrace the idea that the public has a "right to every man's evidence,"[53] they are more circumspect about the right to every man's expert. Despite their ubiquity, experts continue to generate concerns for the judicial system. Doubts have been voiced about overly biased experts, inaccurate conclusions, misleading testimony,

 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

jury incomprehension, and the fear that scientific expert testimony may possess an "aura of infallibility."[54]

This section focuses on a more contemporary concern engendered by the Supreme Court's decision in *General Electric v. Joiner,* where the Court remarked that expert testimony may not be admitted where the gap between the data assembled and the opinion given is too wide. This so-called *ipse dixit* problem[55] occurs with much expert evidence and, remarkably often, with forensic individualization testimony; in addition, problems of outright fraud and other issues that can affect the reliability of forensic evidence must be examined.

The Daubert *trilogy.* Starting in 1993, the U.S. Supreme Court issued a ground-breaking series of decisions governing the use of expert evidence at trial: *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), *General Electric Co. v. Joiner* (1997), and *Kumho Tire Co. Ltd v. Carmichael* (1999).[56]

The Court granted certiorari in *Daubert* to determine whether *Frye*'s general acceptance standard survived the enactment of the Federal Rules of Evidence.[57] Before *Daubert*, the *Frye* test often governed the admission of novel scientific evidence in federal and many state courts. The Court held that *Frye*'s general acceptance test was superseded by the enactment of the rules.[58]

The *Daubert* decision, which governs all, not just novel, scientific evidence, focused on both reliability and relevance. The Court determined that such evidence must be both grounded in the methods and procedures of science and relevant to the precise inquiry at hand ("fit").[59] Moreover, the Court stressed, the focus should be solely on the principles and methodologies employed, not the conclusions generated.[60]

To be considered scientific knowledge, the Court continued, an inference or assertion must be derived by the scientific method, i.e., have appropriate validation. While "not presum[ing] to set out a definitive checklist or test," the Court made a number of "observations" that could be helpful in determining reliability: (1) whether the theory or technique can be or has been tested; (2) whether the theory or technique has been subjected to peer review and publication—an aid to evaluating the quality of the testing, the court explains; (3) the potential or known rate of error of the theory or technique when applied; (4) the existence and maintenance of standards controlling the technique's operation; and (5) whether the technique or theory has been generally accepted in the relevant scientific community.[61]

This flexible standard for expert testimony was premised on the idea of "evidentiary reliability" or "trustworthiness." Rather than wholesale exclusion of evidence, *Daubert* opined that "[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof," coupled with the court's power to direct verdicts and grant summary judgment, were the "appropriate means of attacking shaky but admissible" evidence.[62] The difficulty, of course, is determining where "shaky but admissible" ends and "unreliable thus inadmissible" begins.

The subsequent cases, *Joiner* and *Kumho Tire*, answered questions that developed in the wake of *Daubert*. In *Joiner*, the court softened the line it had earlier drawn between methodology and conclusions, stating:

> conclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.[63]

Kumho Tire, decided in 1999, held that the *Daubert* gatekeeping requirement applied to all expert evidence, not just to scientific evidence, and also reaffirmed the trial judge's latitude in determining what test is appropriate for the type of expertise before the court and whether the proposed testimony meets that test:

> We . . . conclude that a trial court *may* consider one or more of the more specific factors that *Daubert* mentioned when doing so will help determine that testimony's reliability. But . . . the test of reliability is "flexible," and . . . the law grants a district court the same broad latitude when it decides *how* to determine reliability as it enjoys in respect to its ultimate reliability determination.[64]

Although the trilogy embraces a flexible standard, the language of *Kumho Tire* provides an important recommendation: "[A] trial court *should* consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony."[65] When such language is coupled with *Daubert*'s concern for "*evidentiary* reliability—that is, trustworthiness,"[66] it becomes apparent that the Supreme Court is urging courts to admit only trustworthy expert evidence based on sufficient facts and data, and to disallow expert evidence that rests on an inadequate foundation or the *ipse dixit* of the expert. Although the standard of evidentiary reliability is flexible, the language of *Daubert* and *Kumho Tire* provides guidance. As Justice Scalia observes in his concurrence in *Kumho Tire*, while "the *Daubert* factors are

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

not holy writ, in a particular case the failure to apply one or another of them may be unreasonable and hence an abuse of discretion."[67]

Rule 702 of the Federal Rules of Evidence was amended in 2000 to reflect the Supreme Court's requirements for expert evidence, and many state courts have followed the lead of the Court, incorporating a reliability-based standard to govern the admission of expert evidence. Although a number of states still use a general acceptance standard to analyze expert evidence, many of those states also focus on the trustworthiness of such evidence.

Although the *Daubert* factors may make little sense when evaluating the testimony of plumbers in a case of negligent installation of a bathroom sink, they make eminent good sense when evaluating the myriad forms of forensic evidence relevant to the issue of evidentiary reliability.

Problems of forensic identification and individualization. Concerns about the gulf between existing data and the opinions offered have led courts to exclude expert opinions in civil cases, notably in cases involving medical, toxicological, epidemiological, or engineering testimony, where the court has found too great a gap between the data and the opinion expressed.[68]

Problems relevant to evidentiary trustworthiness also are posed by forensic individualization, including the likelihood of error, the inherent problems of declaring a match, the lack of objective standards, and failure to follow scientific method, including blind testing. But early in the history of the forensic identification sciences, courts set aside their traditional caution and freely admitted nearly every specie of such testimony, asking little of the proponents—even to establish the validity of their various claims.[69] In their willingness to be less searching with forensic identification evidence than with other types of scientific evidence, courts opened their doors to a risk of error, which today's courts are only beginning to recognize.

The *Daubert* trilogy has ushered in an era of belated questioning of forensic science. In some instances, forensic science expert testimony that had gone unchallenged for many decades has found itself limited or even excluded from admission. As Paul Giannelli and Edward Imwinkelried have put it: "All the areas of forensic science discussed in this article share two common denominators: In each area 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."[70] Some proponents of admission have argued that the ready acceptance by early courts made many forensic sciences feel that continuing research and self-scrutiny were unnecessary and that, because the low threshold set by courts was responsible for the fields' weak foundations, it would now be "unfair" to the forensic sciences to be confronted with a more demanding test.[71] However unavailing the argument, it does make an important point: the tests posed by courts affect the quality of some fields, especially those whose ultimate and often only audience is the courts. Why exert the effort to produce a product of better quality when the customer is satisfied with one of lower quality?

Errors happen. DNA typing has revealed something quite surprising about the rest of forensic science: how common errors are in forensic science. Because DNA typing has convincingly exonerated numerous persons who had been wrongly convicted, it has created an opportunity to reexamine trial evidence in an effort to discover what led to the erroneous convictions in the first place. One such inquiry found forensic science errors to be second only to eyewitness errors as a cause of wrongful convictions. As Table 1, this page, shows, 66 percent of the cases of wrongful conviction studied had erroneous forensic science presented at trial.

The risks of error have been revealed in more systematic and broad-ranging ways by proficiency studies. Before proficiency studies, forensic scientists could assert that their conclusions rarely, if ever, were in error and that all examiners would agree with what one examiner concluded about a given examination. As long as no evidence existed testing these claims, examiners could say almost anything without risk of contradiction. In most cases, an item of evidence would be examined, an opinion would be announced, no one would know whether the examiner was right or wrong, and no one would know whether or not other examiners might have reached other conclusions.

With the advent of proficiency testing, known materials were for the first time sent to a (sometimes large) number of examiners, who were asked to perform their usual examination of the evidence. Their conclusions could then be compared to the correct answers and to the conclusions of other examiners. These tests revealed the following data: examiners sometimes reached erroneous conclusions; some fields on average made more errors than others; within specialties the level of incorrect conclusions varied depending on the nature of the evidence to be examined (i.e., some tasks are more difficult than others); and for some (if not most) fields, accuracy was not significantly related to years of experience, certification, or other qualifications.

Assertions of near-perfect accuracy had to be tempered in the face of proficiency testing results showing spectrographic voice identification error rates as high as 63 percent; handwriting error rates ranging as high as 100 percent (though the more realistic rate of errors appears to be nearly 40 percent); false posi-

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

tive error rates for bite marks as high as 64 percent; tool-mark identification errors as high as 35 percent; and as many as one-quarter of fingerprint examiners failing to correctly identify all latent prints in a typical proficiency test.[72]

Still, these tests probably understate the actual error rate in everyday casework. The tests often are relatively easy.[73] The tests are non-blind, so that participants are aware that their accuracy is being evaluated, and they know to work harder to try to avoid the embarrassment of making discernible errors. Also, in contrast to everyday cases, proficiency tests provide no extrinsic cues to expected or desired answers, and change the examiners' base rate assumptions.

Of course, if some examiners make errors and others do not, examiners cannot all be in agreement with each other. Thus, claims that all qualified document examiners would reach the same conclusions in all cases,[74] for example, though once asserted confidently, can now be seen to be groundless exaggerations—hopes and assumptions offered as facts.

Proficiency testing might come to serve a far more valuable research purpose. If tests were conducted by independent external organizations, were blind (so that examiners cannot distinguish proficiency tests from casework), and used systematically chosen realistic problems, both examiners and courts would have the benefit of a map of each specialty's strengths and weaknesses, and would know which examination tasks each performs well, less well, or barely adequately.

Clarifying the meaning of a match. A second important thing that DNA typing has done is to provide forensic science and the courts with the first scientifically grounded approach to forensic identification. The model provided by DNA typing serves as a benchmark of quality forensic evidence, illuminates the weaknesses of the other forensic sciences, and provides a guide to future improvement.

Because DNA typing is the first individualization technique to emerge from traditional sciences (namely, molecular biology and population genetics), it refrains from relying on unproven (and likely unprovable) assumptions of uniqueness to reach its conclusions. Thus, DNA typing begins but does not end with a judgment that two samples appear indistinguishably alike—it recognizes that such an observation does not necessarily lead to the conclusion that the two samples came from the same person, because it is both practically and theoretically impossible to know whether all members of any large class are in fact unique. But, even if they are unique, crime scene evidence usually provides examiners with only samples, not full objects. A dozen alleles (alternative forms of a gene), not the entire genome, is what DNA typing looks at; latent fingerprints are usually only partial prints; handwriting is a tiny sample from a vast potential output from any given writer; bite marks are only partial; tools and weapons change with use; and so on. So it is necessary to use methods of analysis that are capable of making useful estimates about *samples* or a temporally changing target.

Table 1

Factors Associated with Erroneous Convictions

Eyewitness errors	74%
Forensic science	
Erroneous	66%
Fraudulent/Misleading	31%
Police misconduct	44%
Prosecutorial misconduct	40%
Bad lawyering	28%
False confessions	19%
Dishonest informants	17%
False witness testimony	17%

Source: M.J. Saks & J. J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892 (2005).

Thus, in DNA typing, examiners calculate the probability of a random match (RMP) in order to obtain an estimate of the likelihood that other persons could match the questioned sample as well as the suspect does. To do this, researchers collect population samples and count the frequency of occurrence of various genetic attributes in the population. To calculate the RMP accurately, the correct probability model must be applied. The most familiar of these is the "product rule," which is appropriate when the attributes are independent of (that is, uncorrelated with) one another.[75] From the RMP, fact finders can gain a proper sense of the rarity of the *perpetrator's* DNA and, therefore, the likelihood that someone *other than* the defendant could have left the DNA found at the crime scene.

Forensic individualization sciences other than DNA typing, by assuming uniqueness, exempt themselves from the following possibilities: collecting data about the distribution of fingerprint or handwriting or bite mark (or whatever) attributes in the population; determining the independence or dependence of these features; finding and applying the correct probability model; calculating RMPs; and sharing the data resulting from the analysis with the court so that fact finders have an informed sense of the likelihood that someone other than the defendant

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

left the questioned evidence. Fields other than DNA typing substitute intuition, anecdotes, or subjective impressions for hard data—or not even that, instead merely leaping from an observation of similarity to a conclusion of common source.

Ironically, all other forensic individualization sciences offer the same rationale that DNA typing does—essentially, the product rule—as the theory underlying their field's ability to link crime scene evidence with suspects. But only DNA typing employs those concepts to perform actual casework and to offer RMPs to fact finders.

Other forensic individualization sciences use DNA typing only as a metaphor to justify their assertion of uniqueness, on which is based the assumption that if two things appear indistinguishably alike, they "must" have originated from a single common source. But a probability, however small, does not lead to an inference of uniqueness. It leads only where DNA examination takes it: to a probability. Serious students of fingerprints long ago realized that probability did not prove uniqueness: "[The probability] approach carries the implication that a complete correspondence of two patterns might occur. . . . It is impossible to offer decisive proof that no two fingerprints bear identical patterns."[76]

"I know one when I see one": The subjectivity problem. From the preceding discussion, one might conclude that determining whether something is a match or not is relatively straightforward and that the difficulty lies only in interpreting the meaning of that match. But even this initial step is fraught with more problems than even many forensic scientists know.

When two examiners look at the same visualization of evidence, they see different things more often than we might expect. We should not be surprised, considering that the examination process is overwhelmingly subjective, with few external standards to guide whether two images should be judged indistinguishably alike or not. In the United States, no minimum number of points of comparison is required to conclude that two fingerprints are indistinguishably alike. "When a fingerprint examiner determines that there is *enough* corresponding detail to warrant the conclusion of absolute identification, then the criteria have been met."[77] In other words, there are no criteria. A British study of 130 fingerprint experts presented with ten pairs of prints found that the examiners often disagreed on how many points of similarity a pair of prints shared. For one of the pairs, the reported number of matching points ranged from ten to forty; for another, it ranged from fourteen to fifty-six.[78] This is not science; it is individual perception and judgment. Although an FBI supervisory fingerprint examiner, a verifier, and the chief of the Latent Print Unit who worked the 2004 Madrid, Spain, bombing case believed they saw a match, other experts later said that the latent and the file print were "filled with dissimilarities."[79]

If this is true of fingerprints, the gold standard of assumed precision, accuracy, and complete consensus, imagine the extent of disagreement among examiners of handwriting, tool marks, bite marks, and so on.

Where are the scientists? "Forensic scientist" is a misleading title. In the world of conventional science, academically gifted students spend at least four years after college in doctoral training, where much of the socialization into the culture of science as well as specialized education takes place. That culture emphasizes obsessive methodological rigor, openness, relentless criticism of methods and findings, and cautious interpretation of data. Science proceeds by testing hypotheses using double-blind, controlled, repeatable studies that are published only after careful review of methods and logic. Published studies are then subjected to further critical review by the wider community of scientists—much of which describes *Daubert*'s view of the road to reliability.

By contrast, those who routinely testify under the appellation of "forensic scientist" operate in a much different world. In the forensic sciences, where 96 percent of practitioners hold bachelor's degrees or less, 3 percent master's degrees, and 1 percent doctoral degrees,[80] it is hard to find the culture of science.

Most forensic "scientists" have little understanding of scientific methodology, do not design or conduct research (and do not know how to), often have not read the serious scientific literature beginning to emerge in their fields (often conducted by doctoral level scientists from other fields), and would be unable to critique these studies sufficiently for the standards of conventional peer-reviewed scientific journals. Scientific findings relevant to a given forensic science often are ignored in the conduct of everyday casework.

A more descriptive title in most instances would be "technician." Yet the nominal upgrading of a title to forensic "scientist" has obvious advantages in the world of litigation, where appearance often can serve as well as reality. If hard questions are asked about the underlying science, typical forensic science witnesses can explain that they cannot be expected to respond to such questions because the work they do, and the training they receive, is really that of a technician. Sometimes a "genuine scientist" (not infrequently from a field different from the forensic discipline at issue) may be brought in to help.

The future is coming, however. Forensic science services in Europe, Australia, and New Zealand employ many more real scientists designing real research to test, study, inform,

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

and improve the work of all crime laboratories. There is also a small but growing cadre of serious American forensic scientists working and conducting research in both academic settings and crime laboratories. Given that they often must swim against a complacent tide, these American forensic scientists deserve considerable praise. But it is clear that the numbers, the culture, the support, the research, the deepest thinking, and the most improved techniques are more often found among scientists working outside of the United States.

The problems of expectation and suggestion. Many fields recognize the risk that people tend to "see what they expect to see." That errors tend to be made in this direction has been well documented in a large body of research conducted by a variety of fields: astronomy, psychology, medicine, and others.[81] To prevent these effects from distorting perceptions, judgments, and conclusions, many fields have structured "blindness" (or "masking") into their procedures. Blind taste tests, blind grading, blind review by peer-reviewed journals, and the blind and double-blind procedures used in much scientific research are familiar examples.

The work of forensic scientists subjects them to many opportunities to be affected by expectation and suggestion. Before conducting their examinations, they sometimes learn facts or theories about the case from police investigators or from transmittal letters or forms accompanying the evidence and requesting the examination. (These often tell the examiner what the investigators expect or hope will result from the examination.) Examiners sometimes learn what fellow examiners testing different items of evidence found, which they then are likely to expect will be consistent with their own examinations.[82] One of the few published audits of crime laboratories found that examinations resulted in conclusions excluding submitted evidence only 10 percent of the time on average—a finding consistent with expectation and suggestion effects.[83]

The problem of unintended bias can be organizational as well as psychological. When examiners produce results that are disappointing to investigators, they sometimes are asked to reexamine only those displeasing results. Rarely, if ever, are they asked to reexamine the results that investigators like. This ensures that only findings inconsistent with an investigator's theory of a case get reversed. All of this can occur with examiners having only the best and most honest of intentions, and yet they can produce conclusions that undermine true facts.

Last year, when FBI fingerprint examiners identified the wrong person as the source of the latent print belonging to one of the terrorists involved in the massive train bombing in Madrid, a special review committee was convened by the FBI to study the error. It concluded:

> The power of the [computer-aided selection of a candidate print], coupled with the inherent pressure of working on an extremely high profile case, was thought to have influenced the examiner's initial judgment and subsequent examination. This influence was recognized as confirmation bias (or context effect) and describes the mind-set in which the expectations with which people approach a task of observation will affect their perceptions and interpretations of what they observe.[84]

The report went on to explain that the first examiner's already heightened expectation of identification was reinforced by finding a small number of matching points, which in turn caused the examiner to mistakenly interpret other points of comparison as matching. Once this "mind-set occurred with the initial examiner, the subsequent examinations were tainted . . . especially because the initial examiner was a highly respected supervisor with many years of experience. . . ." After another examiner was influenced by and concurred with the first, it became almost impossible for anyone else to reach a contrary conclusion.

Despite occasional recognition of the problem, forensic science in the United States has refused to develop procedures to prevent the problem from being able to occur in the first place. In contrast to many other fields of endeavor, forensic scientists have generally insisted that they can avoid falling prey to such influences merely by telling themselves to be immune.

The problem of fraud. In addition to honest errors, DNA exoneration cases revealed that false or misleading testimony was more likely to be offered by forensic science expert witnesses than by any other kind of dissembling witness. (See Table 1, page 25.) These results confirm findings from other sources suggesting that, as a group, forensic scientists offer considerably more than their share of false testimony.[85] Andre Moenssens, a forensic scientist turned law professor, has reported: "All [forensic science] experts are tempted, many times during their careers, to report positive results when their inquiries came up inconclusive, or indeed to report a negative result as positive. . . ."[86]

Fraudulent expert testimony has been a terrible problem for the criminal justice system. Fred Zain of West Virginia,[87] along with numerous other forensic science experts, has testified fraudulently in many cases.[88] Not only are innocent people wrongly convicted on such evidence, but the state must also review those convictions involving the fraudulent testimony to determine which cases must be retried. This process is an expensive and time-consuming endeavor for the government. More importantly, it is a

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

nightmare for both crime victims and those innocent people sent to prison or death row.

The subjectivity of the examination and the structure of the examination process facilitate the problem of fraud. Often the forensic science expert is the only person to have viewed the evidence through a microscope, and the judgment of similarity and inference of common source take place entirely in the head of that person. If other examiners serve as verifiers, they typically know what the first examiner concluded and (as in the case of the FBI's Madrid analysis) find it difficult to see and conclude anything else. If an examiner claims to have drawn an inference of common source between indistinguishably similar questioned and known evidence, with what can one test the candor of such an opinion?

Many of the problems discussed above are evident in the case of *Commonwealth v. Gregory*.[89] Gregory was convicted of rape and attempted rape and sentenced to seventy years in prison. The government's strongest evidence against him was the testimony of a forensic scientist testifying about a microscopic hair examination that, she stated, found the defendant's head hair to be indistinguishable from the hair found in a pantyhose mask worn by the assailant. She stated that the features of the hair were so unusual that it was extremely unlikely that they could have come from any other person. The witness conceded that there were no data on which such inferences could be based, but because of her experience ("when you've examined as many hairs as I have . . .") she knew an identification when she made one. Gregory spent seven years in prison before DNA testing excluded him as the rapist and led to his conviction being vacated.

The defense attorney at this trial assumed the basic honesty of the expert's reported observations, and his questions focused on the basis for her inference of identity. In reviewing the evidence from this DNA exoneration case, most commentators regarded the outcome merely as a typical case of an expert's having too much confidence in her own subjective guesstimation of rarity and offering testimony that exaggerated the inculpatory implications of a hair that was indistinguishably similar. She thereby exceeded what the field claims for its own techniques but explained, as experts seem to do surprisingly often, that this case was "different" from most. The expert's testimony that the hairs appeared identical, however, was never doubted.

Not until the suit for a federal civil damages trial against the city, the police, and the expert was being prepared did another expert—one hired by plaintiff's counsel—examine the hair—and conclude that the hairs did not match at all.[90] The expert had fabricated her report and her testimony to help win a conviction. In the end, this was another case of a forensic scientist who decided to report a negative result as positive.

Judicial Gatekeeping: Developing Workable Filters

The single most important observation about judicial gatekeeping of forensic science is that most judges under most circumstances admit most forensic science. There is almost no expert testimony so threadbare that it will not be admitted if it comes to a criminal proceeding under the banner of forensic science. (Handwriting and voiceprint identification are the chief exceptions to this generalization.) With this background in mind, we will now explore possibilities for improving the judicial management of such testimony. If—for whatever reasons and regardless of the governing rules—judges feel they must admit expert testimony that cannot establish the validity of its claims, can the courts find ways to do so yet still protect fact finders from being misled by unfounded expert exaggerations?

Some forensic sciences have been with us for so long, and judges have developed such faith in them, that they are admitted even if they fail to meet minimal standards under *Daubert*. Faith, not science, has informed this gatekeeping. For example, fingerprint experts have proved to be remarkably short on studies that support their claims. (In the words of one judge, "the government had little success in identifying scientific testing that tended to establish the reliability of fingerprint identifications."[91]) Through dozens of *Daubert* challenges, no judge has been presented with conventional scientific evidence capable of persuading a rational gatekeeper, yet at the same time no court has excluded the opinions of fingerprint examiners. Various judges have found various paths around *Daubert*.[92]

Far more worrisome is that proffers of science unsupported by science have received similar treatment. Maverick experts claim to have techniques that have not been heard of or tested, that are based on nothing but the witness's say-so, that others in the field believe do not exist, and that can be performed only by the maverick. Admitted.[93] The maverick is expelled from his professional organization for scientific fraud, yet still may be offered to courts as an expert witness. Admitted.[94] A whole field has no studies validating or even testing its claims. Admitted. The few studies undertaken by a field suffer from fundamental flaws. Admitted. The studies show high error rates. Admitted.[95]

The applicable legal test offers little assurance. The maverick who is a field unto him- or herself has repeatedly been readily admitted under *Frye*, and the complete absence of foundational research has not prevented such admission in

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

Daubert jurisdictions. As noted, both historical and contemporary exceptions exist, such as asserted handwriting and voiceprint identification, but to little effect.

How, then, can judges admit forensic science testimony while still protecting juries from an assemblage of witnesses who are second only to eyewitnesses in their presentation of errors and exaggerations—and in fraud are second to none? Courts can control the gate governing the admission of testimony in several ways: partial or limited admission of evidence, renewed focus on the task at hand, limiting or disallowing overpowering and misleading terminology, allowing competing opinions into evidence, and providing limiting jury instructions. Each of these is briefly discussed below.

Limited admissibility and disallowing conclusions of a match. The Federal Rules of Evidence recognize that one basis of expertise is experience and training. Clearly, those examiners who compare either fingerprints or handwriting have a basis of knowledge premised upon repetition of, and familiarity with, the task. Although any opinion about a conclusive match is no more than *ipse dixit*, the expert is not necessarily precluded from pointing out similarities and differences that a lay jury might otherwise not recognize. It is in the *interpretation* of those similarities that examiners are on the shakiest ground. Thus, the most expedient "fix" to the problem of missing data supporting these individualization specialties is to allow the examiner to discuss the points of comparison but to disallow the examiner from declaring a match or asserting conclusions.

Various federal district courts have taken such an approach, particularly with handwriting comparison.[96] These courts recognize that examiners may be trained to identify points of similarity and dissimilarity between samples, and thus may be more perceptive than jurors. These courts expect descriptive testimony to be helpful to the jury but at the same time recognize that with no data from which to extrapolate a probability of the likelihood of a random match, there is no basis for allowing opinion testimony concerning an identification.

Focusing on the task at hand. In both *Daubert* and *Kumho Tire*, the Supreme Court noted that the testimony must be both reliable and "relevant to the task at hand."[97] The proposed expertise must be an accurate fit with the issue on which an expert will testify. The inquiry is whether the "expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute."[98]

Although courts and litigants often focus primarily on the reliability factors, it is likewise important for courts to determine whether the expertise claimed is sufficiently close to the issue in dispute. In considering the admissibility of handwriting expert testimony, for example, a district court disallowed an expert from testifying about hand printing (as opposed to cursive script) on an immigration form alleged to have been filled out by the defendant.[99] The court found that there was insufficient evidence whether the comparison of hand printing was a valid expertise.[100] Moreover, the court found that since the defendant was raised in Japan where he had learned to print English at school, there was no evidence that the examiner could recognize subtle similarities and differences of Japanese native hand printers.[101] Thus, although the expert might be able to form sound inferences about authorship of cursive writings of a native English writer (on which the court expressed no opinion), the witness was *not* an expert on the task at hand, nor was there sufficient evidence to establish that the task at hand was even an actual area of expertise.

In a sense, the focus on task at hand is an inquiry related to decisions about qualifications. For example, a physician may be qualified to testify about a variety of medical issues, but she may not be qualified to testify about particular medical controversies at issue in a given case—a finding courts have made in medical device litigation.[102]

Thus, if a court determines that there is insufficient fit between the proffered testimony and the precise inquiry before the court, the court may disallow the testimony entirely or cabin the testimony within the bounds of the established expertise.

Disallowing overpowering or misleading testimony. Some terminology that experts use implies a far stronger probability than is intended or justified by the underlying science.

At least one field insists upon using terminology that is absolute, categorically concluding that an "identification" exists or does not (or that no opinion can be reached). The field of fingerprint examination goes so far as to prohibit its members from offering qualified opinions (which amounts to anything less than these categorical alternatives). Ironically, only a probabilistic statement—and a subjective probability estimate at that—could accurately reflect the real basis of the judgment the examiner has made. There is no basis for absolute opinions.[103]

Some fields use terms that do not readily convey to laypersons what the examiner intends. For example, if a forensic dentist states that in his opinion a "match" has been found between the bite mark in a victim and the defendant's dentition, the jury is likely to infer that means that the defendant is the biter. But, for forensic dentists, the actual meaning of a match is: "some concordance,

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

some similarity, but no expression of specificity intended; generally similar but true for a large percentage of population."[104] Other fields' experts might use the word "match" to mean a very strong probability of common source, but the judge or jury may well think it means an even more certain linkage than is intended. A "match" merely indicates that two marks or objects appear indistinguishably alike. A second step must be performed to interpret the likelihood that two things that "match" originated from a single source.

One of the authors of this article has presented some of the recommended or official terms of a variety of forensic identification disciplines to law students and to college students, asking them to indicate on a scale from 0 to 100 what they thought the expert was seeking to express about the likelihood that the questioned evidence came from the defendant. Their numerical estimates often were far from what the official definitions intended, often leading to inversions of meaning. Often, terms meant to express very strong linkage—such as "reasonable scientific certainty"—were taken to mean much weaker linkage, and vice versa.

Sometimes expert witnesses and fields have developed terms that are conceived to be defensible as literal truth but are likely to induce the jury to think a more inculpatory opinion has been offered than the witness might assert given the evidence. "Consistent with" is one such example. Some examiners have adopted more blatant phraseology; "indeed and without doubt" is an example.[105]

Witnesses should be confined to meaningful language that accurately conveys inferences that genuinely can be supported by the field's methods. Overreaching and exaggeration should be banned from the witness box. Any expression of absolute certainty by forensic identification experts, or any term likely to be understood by the fact finder as conveying such a strong and unjustifiable meaning, should be prohibited. Among such suspect terms would be "identification," "match," "unique," "no other in the world," "identification to the exclusion of all others in the world," and "consistent with."

One of the most straightforward ways of tempering implied or explicit exaggeration is to require experts to accurately inform the jury about error rates. What could matter more to reliability and weight than error rates? How often do experts in a field reach the correct or incorrect conclusions *in the task relevant to the case at bar*? If there is a lack of data on the precise question at issue, or from the field more generally, that itself is informative.

Use of court-appointed experts and panels of experts. The law does not limit judges to hearing from only the experts offered by the parties. In appropriate circumstances, such as deciding whether or how much evidence to admit, judges have the power to obtain additional expertise. In various jurisdictions, by rule or by common-law power, courts may appoint their own expert witnesses, panels of experts, advisory juries, or consulting experts. Selecting and communicating with such experts must, of course, be handled with care.

The power of courts to appoint their own expert witnesses has been codified in Federal Rule of Evidence 706 and affirmed by the Supreme Court in *Daubert*, where the Court recognized a trial court's ability to "procure the assistance of an expert of its own choosing."[106] In making admissibility decisions, some courts have appointed consulting experts who do not testify but in effect sit at the judge's side to help the judge understand the scientific testimony and arguments that are offered at the Rule 104(a) hearing. It is within Rule 706 or the inherent authority of the court under the Evidence Rules that the power of such an appointment has been found.[107]

Science panels have been appointed to assist the court in complex tort cases, notably in the silicone breast implant litigation.[108] Given the myriad forms of forensic science and the number of challenges raised to such evidence, these panels could be quite helpful to courts deciding the admissibility of controversial forensic science.

Appointment and allowance of competing expert opinions. Defendants typically come to court without experts, whether to challenge admissibility in limine or to testify at trial to challenge the accuracy or weight of the government's testimony. The most obvious reason for this is a lack of funds to pay for them. In addition, many criminal defense lawyers do not anticipate needing an expert, assuming all forensic evidence will be reliable or they can expose all of the weaknesses they need to on cross-examination. Though both federal and state courts provide for appointment of expert witnesses, many jurisdictions have high standards for appointing experts.

Greater access to and use of competing expert evidence in criminal cases involving forensic identification testimony would dovetail with *Daubert*'s recognition that "presentation of contrary evidence" is one appropriate method of attacking shaky but admissible evidence.[109] For the adversary process to work, both advocates need the resources to present their strongest case. Courts can help make the adversary process work as it should by making such appointments more regularly. In the long run, it would also lead to improvements in forensic science because, ironically, the weaknesses in forensic science are in considerable part the result of a lack of adversarial testing of forensic science throughout most of the twentieth century.

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

Jury instructions. Some courts have decided that if shaky but admissible expert forensic evidence comes before the jury, the jury should be instructed in a manner that allows it to put the evidence in proper context. The court explains to the jury that forensic scientists are not scientists per se but are more akin to craftsmen, and that their testimony may be less precise than, for example, a chemist's.[110] The concern that jurors may overvalue expert opinion evidence can be ameliorated somewhat by a clear instruction designed to temper the belief jurors may have in the aura of infallibility of forensic examiners.

Conclusion

As Justice Stephen Breyer noted in his concurrence in *General Electric v. Joiner*, scientific evidence often requires judges to make "subtle but sophisticated" determinations; at the place where "law and science intersect, those duties often must be exercised with special care."[111] Make no mistake, these are not simple tasks. In an attempt to help judges in those difficult determinations, we have sought to provide some guidance in the area of forensic science, offering an approach that conforms to the law and yet recognizes the realities of judicial decision making in a complex area.

Endnotes

1. Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1316 (9th Cir. 1995), on remand from the Supreme Court decision in Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).

2. *See* Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 149 (1999) (extending the "evidentiary reliability" requirement to all expert testimony). Together with General Electric Co. v. Joiner, 522 U.S. 136 (1997), and the Supreme Court decision in *Daubert*, these cases are collectively often referred to as the "trilogy on expert evidence" or "the *Daubert* trilogy." For a state-by-state review of standards governing expert testimony, *see* JANE CAMPBELL MORIARTY, PSYCHOLOGICAL AND SCIENTIFIC EVIDENCE IN CRIMINAL TRIALS, app. 1 (1996, 2005 Supp.).

3. For further discussion of this distinction, *see* John I. Thornton & Joseph L. Peterson, *The General Assumptions and Rationale of Forensic Identification*, §§ 31:9-31:11, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 2005).

4. *See* MORIARTY, *supra* note 2, § 12:4, at 12-6 through 12-9 (discussing fraud and negligence at various crime labs).

5. To complicate matters, the word "individualization" often is used interchangeably with "identify," so that the terminology easily masks important differences between identification and individualization. *See* Thornton & Peterson, *supra* note 3.

6. R.E. Gaensslen, *Forensic Analysis of Biological Evidence, in* 3 FORENSIC SCIENCES § 29.04[b], at 29-11 (Cyril H. Wecht ed., 2000).

7. *Id.* at § 29.03[b], at 29-9.

8. *Id.* at § 29.06, at 29-20.

9. PAUL C. GIANNELLI & EDWARD J. IMWINKELRIED, SCIENTIFIC EVIDENCE § 23-2(D), at 362 (3d ed. 1999).

10. *Id.* at 366.

11. ANDRE A. MOENSSENS ET AL., SCIENTIFIC EVIDENCE IN CIVIL AND CRIMINAL CASES § 9.05, at 560 (4th ed. 1995).

12. *Id.* § 9.05 at 561. *Accord* Georg Jochem, *Fiber Plastic Fusions and Related Trace Material in Traffic Accident Investigation*, in TRACE EVIDENCE ANALYSIS 97 (Max M. Houck ed., 2004).

13. GIANNELLI & IMWINKELRIED, *supra* note 9, § 25-2, at 512, § 24-2(L), at 446.

14. *Id.*, § 25-2(A), at 515–16.

15. Michael J. Saks, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 HASTINGS L.J. 1069 (1998).

16. For further reading, *see* Thornton & Peterson, *supra* note 3, ch. 31.

17. MORIARTY, *supra* note 2, § 11:57, at 11-116, 11-117

18. *See* Daniel Krane & William C. Thompson, *DNA Evidence* §§ 11:14 to 11:23, at 11-22 through 11-37, *in* PSYCHOLOGICAL AND SCIENTIFIC EVIDENCE IN CRIMINAL TRIALS (Jane Campbell Moriarty ed., 1996, 2005 Supp.); David H. Kaye & George F. Sensabaugh Jr., *The Scientific Status of DNA Typing* §§ 32:21-32:60, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 2005) (explaining the scientific foundation of DNA profiling).

19. Paul C. Giannelli, *Mitochondrial DNA*, 19 CRIM. JUST. 54, 54 (Winter 2005).

20. *See* Marlan D. Walker, *Mitochondrial DNA Evidence in* State v. Pappas, 43 JURIMETRICS J. 427, 429 (2003); Edward K. Cheng, *Mitochondrial DNA: Emerging Legal Issues*, 13 J.L. & POL'Y 99, 100 (2005).

21. Cheng, *supra* note 20, at 105. For a detailed but comprehensible explanation of how mtDNA is extracted and the differences between mtDNA and nDNA, see *State v. Pappas*, 776 A.2d 1091, 1101–04 (Conn. 2001).

22. *See, e.g.*, State v. Underwood, 518 S.E.2d 231, 240 (N.C. App. 1999) (holding the evidence reliable); State v. Council, 515 S.E.2d 508, 518 (S.C. 1999) (same); *Pappas*, 776 A.2d 1091 (same). *See also* Wagner v. State, 864 A.2d 1037, 1046-47 (Md. App. 2005) (finding mtDNA testing generally accepted as reliable and collecting other cases with similar findings).

23. *See, e.g.*, People v. Jennings, 96 N.E. 1077, 1081 (Ill. 1911) (holding fingerprints admissible).

24. *See* Simon A. Cole, *Fingerprint Evidence* § 12-14, at 12-13, *in* PSYCHOLOGICAL AND SCIENTIFIC EVIDENCE IN CRIMINAL TRIALS (Jane Campbell Moriarty ed., 2005 Supp.). For a complete description of the process and its limitations, *see* David A. Stoney, *The Scientific Basis of Expert Testimony on Fingerprint Identification* §§ 34:21-34:53, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 2005).

25. Stoney, *supra* note 24, at §§ 34:21-34:26.

26. *Id.* at § 34:28.

27. Cole, *supra* note 24, § 12:18, at 12-16.

28. Stoney, *supra* note 24, at § 34:28.

29. Cole, *supra* note 24, § 12:20, at 12-19.

30. *Id.*

31. *See European Fingerprint Standards*, 28 FINGERPRINT WHORLD 19 (2002) (reporting this point range among European countries), *cited in* Cole, *supra* note 24, § 12:22, at 12-22.

32. Cole, *supra* note 24, § 12:23, at 12-23.

33. *See* Alfred Biasotti & John Murdock, *The Scientific Basis of Firearms and Toolmark Identification*, § 36:6, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 2005); Adina Schwartz, *Toolmark and Firearm Identification* § 12:39, at 12-49, *in* PSYCHOLOGICAL AND SCIENTIFIC EVIDENCE IN CRIMINAL TRIALS (Jane Campbell Moriarty ed., 2005 Supp.).

34. Schwartz, *supra* note 33, § 12:40, at 12-50.

35. Biasotti & Murdock, *supra* note 33, § 36:8.

36. *Id.*

37. *See, e.g.*, MOENSSENS ET AL., *supra* note 11, § 9.13, at 577 ("no technique known today that can positively identify a crime scene hair as having come from a specific individual").

38. Of the first seventy-one convictions

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

that were reversed on the basis of DNA testing, twenty-one involved erroneous microscopic identification of hair samples. Note, however, that error rates are higher for a number of other forensic identification fields.

39. FBI HAIR COMPARISON REPORT (1984), *cited in* MOENSSENS ET AL., *supra* note 11, § 9.10, at 569.

40. GIANNELLI & IMWINKELRIED, *supra* note 9, § 24-2(B)–(C), at 435–38.

41. Some microscopic hair comparison examiners on some occasions testify beyond their field's stated capabilities.

42. *See* Max M. Houck & Bruce Budowle, *Correspondence of Microscopic and Mitochondrial DNA Hair Comparison*, 47 J. FORENSIC SCI. 964, 966 (2002) (finding of all declared microscopic hair identifications, 12 percent in error (as determined by mtDNA testing) and, by implication, of all true exclusions, 35 percent erroneously declared to be identifications).

43. The history is discussed at length in D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise,"* 137 U. PA. L. REV. 731 (1989).

44. Some handwriting examiners, however, testify beyond their field's stated capabilities and offer the opinion that a defendant is the author of a particular forgery.

45. For an in-depth analysis of the problems of reliability and validity associated with handwriting comparison, see D. Michael Risinger, *Handwriting Identification* §§ 35:14–35:37, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 2005).

46. For further discussion of the scientific basis of bite mark identification, see C. Michael Bowers, *The Scientific Status of Bitemark Comparison* §§ 38:8-38:33, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 2005).

47. WILLIAM BODZIAK, FOOTWEAR IMPRESSION EVIDENCE (1999).

48. *See generally* MOENSSENS ET AL., *supra* note 11, § 17.11, at 1049–52.

49. But query whether anything of significance is required for a specialty to go from unrecognized to recognized.

50. COMMITTEE ON SCIENTIFIC ASSESSMENT OF BULLET LEAD ELEMENTAL COMPOSITION COMPARISON, NATIONAL ACADEMY OF SCIENCES, FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE (2004). *See also* United States v. Mikos, 2003 WL 22922197 (N.D. Ill. 2003) (holding source conclusions about CABL inadmissible and explaining with exceptional clarity shortcomings of field's claims).

51. COMMITTEE ON SCIENTIFIC ASSESSMENT OF BULLET LEAD ELEMENTAL COMPOSITION COMPARISON, NATIONAL RESEARCH COUNCIL, FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE (REPORT IN BRIEF) (2004) at Box 1.

52. Charles Pillar, *FBI Ends Controversial Bullet Testing*, L.A. TIMES, Sept. 2, 2005, WLNR 13787541.

53. United States v. Bryan, 339 U.S. 323, 331 (1950), *quoting* WIGMORE, EVIDENCE (3d ed.) § 2192.

54. *See, e.g.*, DAVID H. KAYE, DAVID E. BERNSTEIN, & JENNIFER L. MNOOKIN, THE NEW WIGMORE: A TREATISE ON EVIDENCE, EXPERT EVIDENCE § 10.2, at 334 (2004) (listing various concerns and criticisms of expert testimony).

55. *Ipse dixit* is defined as "an unsupported assertion, usually by a person of standing." AMERICAN HERITAGE DICTIONARY (4th ed. 2000).

56. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993); General Electric Co. v. Joiner, 522 U.S. 136 (1997); Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137 (1999).

57. Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923) (holding proposed evidence "must be sufficiently established to have gained general acceptance in the particular field in which it belongs").

58. *Daubert*, 509 U.S. at 587.

59. *Id.* at 591. Judge Becker coined the concept of "fit" in *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985), a seminal case allowing expert testimony about the vagaries of eyewitness identification.

60. *Daubert*, 509 U.S. at 595.

61. *Id.* at 593–94.

62. *Id.* at 596.

63. General Electric Co. v. Joiner, 522 U.S. 136, 146 (1997). The Court also held that abuse of discretion was the appropriate standard of review. *Id.* at 143.

64. Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 141–42 (1999).

65. *Id.* at 152 (emphasis supplied).

66. *Daubert*, 509 U.S. at 590 n.9 (discussing the distinction between validity and reliability).

67. *Kumho Tire*, 526 U.S. at 159 (Scalia, J., concurring).

68. *See, e.g.*, McLean v. Metabolife Int'l, Inc., 401 F.3d 1233, 1239-52 (11th Cir. 2005) (expert substituted own opinion for scientific proof that use of defendant's supplements caused plaintiffs' strokes and heart attack); Norris v. Baxter Healthcare Corp., 397 F.3d 878, 886 (10th Cir. 2005) (plaintiff experts' opinions that silicone breast implants caused autoimmune diseases not reliably grounded on existing data); *Kumho Tire*, 526 U.S. at 137 (no abuse of discretion in trial court's finding that tire analyst's opinion failed to satisfy reliability criteria).

69. See the legal introductions to chapters on identification by handwriting, tool marks and firearms, fingerprints, bite marks, and voice in DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (2005).

70. Paul Giannelli & Edward Imwinkelried, *Scientific Evidence: The Fallout from the Supreme Court's Decision in* Kumho Tire, 14 CRIM. JUST. 12, 20 (2000).

71. Jamie Orenstein, *Effect of the* Daubert *Decision on Document Examinations from the Prosecutor's Perspective*, FORENSIC SCI. COMM., No. 3, Pt. 5 (Oct. 1999).

72. Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892 (2005).

73. Test participants reportedly complained to test manufacturers when presented with difficult tests. *Id. See also* United States v. Llera Plaza II, 188 F. Supp. 2d 549, 565 (E.D. Pa. 2002) (concluding proficiency testing of FBI fingerprint examiners is less demanding than it should be).

74. D. FISHER, HARD EVIDENCE (1995).

75. Many judges will remember the misuse of the product rule at trial in *People v. Collins*, 68 Cal. 2d 319 (Cal. 1968), where the prosecution expert calculated the probability that an interracial couple in a yellow automobile were persons other than the accused as being "about one in a billion." The correct use of the product rule requires both accurate input probabilities and independence of the attributes from each other. In *Collins*, the expert had no foundation to establish the accuracy of the probability of each attribute, such as a "girl with a ponytail" (he suggested a probability of one in ten). He also had no proof of the independence of the factors, such as black man with a beard and interracial couple in an automobile.

76. H. CUMMINS & C. MIDLO, FINGER PRINTS, PALMS AND SOLES: AN INTRODUCTION TO DERMATOGLYPHICS (1943). *See also* David Stoney, *What Made Us Ever Think We Could Individualize Using Statistics?*, 31 J. FORENSIC SCI. 197 (1991).

77. Stoney, *supra* note 24, § 34:32.

78. Ian W. Evett & R.L. Williams, *A Review of the Sixteen Point Fingerprint Standard in England and Wales*, 46 J. FORENSIC IDENTIFICATION 49 (1996).

79. Robert B. Stacey, *Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 54 J. FORENSIC IDENTIFICATION 706 (2004).

80. Kenneth G. Furton et al., *What Educational Background Do Laboratory Directors Require from Applicants?* 44 J. FORENSIC SCI. 128 (1999).

81. For a review of the literature, see D. Michael Risinger et al., *The* Daubert/Kumho *Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1 (2002).

82. For a discussion of this problem in

 "Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping" by Jane Campbell Moriarty and Michael J. Saks, published in *Judges' Journal*, Volume 44, No.4, Fall 2005 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

forensic science and some suggested resolutions, see *id.*

83. Joseph L. Peterson et al., Forensic Evidence and the Police (National Institute of Justice 1984) (finding average, fewer than 10 percent of reports disassociated suspect from crime scene evidence).

84. Stacey, *supra* note 79, at 713.

85. Paul C. Giannelli, *The Abuse of Scientific Evidence in Criminal Cases: The Need for Independent Crime Laboratories,* 14 Va. J. Soc. Pol'y & L. 439 (1997); Office of the Inspector General, The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases (U.S. Dep't of Justice 1997); *In re* Investigation of W. Va. State Police Crime Lab, Serology Div., 438 S.E.2d 501, 503 (1993).

86. Andre A. Moenssens, *Novel Scientific Evidence in Civil and Criminal Cases: Some Words of Caution,* 84 J. Crim. L. & Criminology 1, 17 (1993).

87. *In re Investigation of W. Va. State Police Crime Lab*, 438 S.E.2d at 501.

88. See collected instances in Craig M. Cooley, *Reforming the Forensic Science Community to Avert the Ultimate Injustice*, 15 Stan. L. & Pol'y Rev. 381, 399–416 (2004). *See also* Moriarty, *supra* note 2, § 12:4, at 12-4, discussing the FBI laboratory errors set forth in the 1997 Justice Department's Office of Inspector General report.

89. The facts are discussed in *Gregory v. City of Louisville et al*., 2004 U.S. Dist. Lexis 7046 (2004).

90. *Id.*

91. United States v. Llera Plaza I, 2002 WL 27305, at *11, *vacated and superceded on reconsideration*, United States v. Llera Plaza II, 188 F. Supp. 2d 549 (E.D. Pa. 2002).

92. David L. Faigman et al. eds., *Fingerprint Identification*, *in* Modern Scientific Evidence: The Law and Science of Expert Testimony, §§ 34:3-34:13 (2005).

93. *See* Giannelli, *supra* note 85 (describing abuses by forensic dentist, forensic anthropologist, and others).

94. *Id.* (describing example of Michael West and others).

95. *See* Faigman et al., *supra* note 69.

96. *See, e.g.*, United States v. Hines, 55 F. Supp. 2d 62, 73 (D. Mass. 1999) (testimony of government's handwriting expert limited to identifying similarities and dissimilarities of known and unknown writing); United States v. McVeigh, 1997 WL 47724 (D. Colo. trans. Feb. 5, 1997) (same); United States v. Van Wyck, 83 F. Supp. 2d 515, 525 (N.D.N.J. 2000) (same); United States v. Rutherford, 104 F. Supp. 2d 1190, 1194 (D. Neb. 2000) (same); Wolf v. Ramsey, 253 F. Supp. 2d 1323, 1347–48 (N.D. Ga. 2003) (disallowing forensic document examiner to give conclusions, noting expert's "explanation for his conclusion seems to be little more than 'Trust me; I'm an expert.'").

97. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 597 (1993); Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 141 (1999).

98. *Daubert*, 509 U.S. at 591, *citing* United States v. Downing, 753 F.2d 1224, 1242 (3d Cir. 1985).

99. *See* United States v. Fujii, 152 F. Supp. 2d 939, 942 (N.D. Ill. 2000). For a thorough discussion about the importance of the task at hand inquiry and about this case, see D. Michael Risinger, *Defining the "Task at Hand": Nonscience Forensic Science After* Kumho Tire Co. v. Carmichael, 57 Wash. & Lee L. Rev. 767 (2000).

100. *Fujii*, 152 F. Supp. 2d at 941.

101. *Id.* at 942 (questioning whether examiner would "have any expertise which would allow her to distinguish between unique characteristics of an individual Japanese hand printer and characteristics that might be common to many or all native Japanese hand printers" and concluding she did not).

102. *See, e.g*., Prohaska v. Sofamor, S.N.C., 138 F. Supp. 2d 422, 438 (W.D.N.Y. 2001); Alexander v. Smith & Nephew, P.L.C., 98 F. Supp. 2d 1310 (N.D. Okla. 2000).

103. Stephen G. Bunch, *Consecutive Matching Striation Criteria: A General Critique*, 45 J. Forensic Sci. 955 (2000).

104. American Board of Forensic Odontology Guidelines for Report Writing Including Terminology, quoted in Bowers, *supra* note 46, at app. D.

105. Giannelli, *supra* note 85 (quoting courts quoting forensic odontologist Michael West).

106. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 595 (1993).

107. *See* Faigman et al., *supra* note 69, § 1:38; Joe S. Cecil & Thomas E. Willging, *Accepting* Daubert*'s Invitation: Defining a Role for Court-Appointed Experts in Assessing Scientific Validity,* 43 Emory L.J. 995, 1002-03 (1994) (discussing law pertaining to advisory experts).

108. *See, e.g.*, Norris v. Baxter Healthcare Corp., 397 F.3d 878, 881-82 (10th Cir. 2005) (discussing appointment and findings of panel).

109. *Daubert*, 509 U.S. at 586.

110. *See, e.g.*, United States v. Starzecpyzel, 880 F. Supp. 1027, 1050–51, app. I (S.D.N.Y. 1995) (providing a draft jury instruction).

111. General Electric Co. v. Joiner, 522 U.S. 136, 147-48 (1997) (Breyer, J., concurring).

 © 2005 by the American Bar Association. Reproduced by permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.