Maine Law Review

Volume 56
Number 1 *SYMPOSIUM: Topics in Law and Technology*

Article 5

January 2004

Scientific Evidence and Forensic Science Since Daubert: Maine Decides to Sit out on the Dance

Thomas L. Bohan

SCIENTIFIC EVIDENCE AND FORENSIC SCIENCE SINCE *DAUBERT*: MAINE DECIDES TO SIT OUT THE DANCE

Thomas L. Bohan

SCIENTIFIC EVIDENCE AND FORENSIC SCIENCE SINCE *DAUBERT*: MAINE DECIDES TO SIT OUT THE DANCE

*Thomas L. Bohan**

I. INTRODUCTION

In 1993, the Supreme Court of the United States stated that with the federal adoption of statutory rules of evidence in 1975, the common law rule for determining admissibility of scientific testimony was superseded, and that thenceforth admissibility of scientific testimony was to be determined solely by Federal Rule of Evidence 702 (Rule 702).[1] At the time, Rule 702 read as follows:

> *Testimony by Experts*—If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.[2]

The *Frye* standard[3] had been adopted in one form or another by most of the federal circuits and by many of the state courts during the 70 years preceding *Daubert*.[4] Referred to as the "general acceptance" standard, the *Frye* standard—although adopted in a variety of forms—had the core requirement that proffered scientific testimony be based on *something* enjoying "general acceptance" among some set of scientists.[5] It was an effort to ensure that expert testimony had some measure of reliability. The *Daubert* Court, in agreement with Petitioners and with the authors of six of the twenty-two amicus briefs[6] that had been filed, held that the strictness of the *Frye* "general acceptance" requirement was not in keeping

---

* Of counsel, Bohan, Mathers & Associates. B.S. Physics 1960 University of Chicago; Ph.D. Physics 1968 University of Illinois-Urbana; J.D. 1980 Franklin Pierce Law Center. Member of Maine and Massachusetts bars since 1980, as well as licensed to practice before the United States Patent and Trademark Office since 1980.

1. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 587-88 (1993).

2. In 2001, Rule 702 was amended to incorporate the gloss put on it by *Daubert*, now reading: If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, *if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.* FED. R. EVID. 702 (emphasis added).

3. Frye v. United States, 293 F. 1013 (App. D.C. 1923).

4. *See, e.g.*, United States v. Tranowski, 659 F.2d 750 (7th. Cir. 1981); United States v. Kilgus, 571 F.2d 508, 510 (9th. Cir. 1978); United States v. Alexander, 526 F.2d 161 (8th. Cir. 1975).

5. Frye v. United States, 293 F. 1013 at 1014. The *Frye* court required that the "scientific principle or discovery" on which the testimony was to be based be "sufficiently established to have gained general acceptance in the particular field in which it belongs." *Id.*

6. Of the rest, fourteen supported Respondent Merrell Dow, arguing in part for a stricter standard for scientific evidence or for the retention of the *Frye* standard, and two supported neither party.

with the goal of the Federal Rules of Evidence to liberalize admission criteria.[7] If the Court had stopped there, *Daubert* would be tantamount to the Maine scientific evidence rule, as set out in *State v. Williams*.[8] But it did not stop there. Instead, engaging in what some might characterize as an exegesis, the Court asserted that since inaccurate expert testimony could not "determine a fact in issue," it was necessary for the trial judge to exclude expert testimony not based on the scientific method.[9] The Court thereby brought in through the back door the same reliability concern that had led to the widespread adoption of *Frye* in the first place. This author asserts that, contrary to popular legal and lay belief, the significance of *Daubert* lies not in its discarding of *Frye* and its emphasis on Rule 702, but rather in its exhorting of trial judges to exercise their "gatekeeper" role with respect to scientific evidence, something that many had been fairly lax about previously.

It is precisely because trial judges have taken this gatekeeper role more seriously than in the past that a revolution is occurring in scientific evidence and forensic science. Adding to the pressure for reexamination and change has been the plethora of DNA-based wrongful-conviction discoveries of the past decade. Men convicted of the most heinous crimes, and often sentenced to death, have subsequently been found indisputably innocent of those crimes.[10]

Just as an autopsy provides a post-mortem check of a physician's cause-of-death finding and/or an earlier diagnosis of disease, the post-conviction DNA analysis can provide a check on the correctness of a verdict or plea.[11] Of course, there is less symmetry in the legal selection process than there is in the medical. Although autopsies are generally sought whenever there is uncertainty in the diagnosis or cause of death, post-conviction DNA reviews are sought only to prove that the guilty verdict was mistaken, that it represented a "false positive." No one seeks such reviews to support a verdict of innocence. Furthermore, it is unlikely that a prosecutor would ever seek such a review to support a verdict (or plea) of guilty.

---

7. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. at 588-90.

8. 388 A.2d 500, 503-04 (Me. 1978).

9. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. at 590-91.

10. Urging this revolution on is the continuing revelation of wrongful convictions—with a total number yet unknown, but 138 at the latest count maintained by The Innocence Project at the Benjamin N. Cardozo School of Law—involving the most heinous crimes, which resulted in people being incarcerated for decades, often under sentence of death. The Innocence Project, *at* http://www.innocenceproject.org (last visited Aug. 26, 2003). In September 2003, Calvin Willis was the most recent prisoner to be freed, after serving twenty-two years of a life-without-parole sentence in a Louisiana prison, wrongly convicted of raping a young child. Keri Kirby, *If it Weren't for DNA, I'd Still be Sitting in Angola*, SHREVEPORT TIMES, Sept. 20, 2003, at A1. An earlier exoneration is described by the Center on Wrongful Convictions of the Northwestern University School of Law. There, a man was freed in mid-1999 after serving fourteen years under sentence of death for a rape and murder that DNA analysis showed was committed by someone else. *Center on Wrongful Convictions, at* http://www.law.northwestern.edu/depts/clinic/wrongful/exonerations/jones.htm (last visited Nov. 22, 2003). *See also* Steve Mills and Ken Armstrong, *Yet Another Death Row Inmate Cleared*, CHI. TRIB. May 18, 1999 at 1.

11. For obvious reasons, post-conviction DNA tests are only useful in a small number of cases. In addition to being limited to situations where DNA from the crime scene was collected and preserved, there has to be a logical nexus between the possible outcome of the analysis and the guilt of the person charged. Typically, post-conviction DNA tests are relevant in crimes involving rape.

Once the wrongful-conviction findings began to surface, there was great interest in investigating what had gone wrong at the underlying trials. It was realized that, in addition to answering the pressing specific question, the results of such an investigation might have significance for criminal trials in general, regardless of the crime charged, and for civil trials. Presumably, errors that were occurring in trials that could be checked with DNA analysis were also occurring in trials for which DNA analysis was not available.[12]

The most common threads running through the trials that led to wrongful convictions are a paucity of evidence and the failure of the defense to put on a forensic expert.[13] In many of the cases, there was no physical evidence at all and the prosecution's case rested entirely on eye witness testimony, sometimes from a single witness.[14] The forensic science community was most dismayed by those cases where the wrongful verdict was based on specious forensic testimony. In most instances, the testimony involved exaggerating, either through implication or direct lying, the significance of those tests that had been done. A typical example would involve the claim that hairs can be "individualized" by microscopic examination, leading to the conclusion that specimens of the defendant's hair had been found at the crime scene. Although the falsity of such statements has long been recognized in professional scientific literature, it seems not to be recognized by the majority of the public.[15] This means that, in the absence of effective opposition, a jury will probably accept the false testimony at face value and as persuasive evidence. Even if the witness only makes a literally true statement that the hair specimen found at the scene "is consistent with" being the defendant's, a jury and judge not familiar with this type of evidence, and not alert to the "is consistent with" subterfuge, can be influenced to the severe detriment of the defendant. A knowledgeable defense expert can help cure such testimony or even prevent it from being offered in the first place.[16]

Section II of this article will go into greater detail with *Frye*, and with the entire *Daubert* trilogy, which includes, in addition to *Daubert* itself, *GE v. Joiner*,[17]

12. Arson cases are a prime example of this. Although forensic investigation techniques underlying arson indictments are not going to be addressed in this article, it is worth noting that long-accepted practices in that field are overdue for validation studies and *Daubert* reviews. It is submitted that if wrongful convictions could be proved as definitively in arson cases as they can in cases involving rape and/or murder, these reviews would have already occurred.

13. *See Advancing Justice through Forensic DNA Technology: Oversight Hearing Before the House Subcommittee on Crime, Terrorism, and Homeland Security, Committee on the Judiciary*, 108th Cong. (2003) (testimony of Peter Neufeld, The Innocence Project), *available at* http://www.house.gov/judiciary/neufeld071703.htm (last visited Oct. 5, 2003). *See also* The Innocence Project, *at* http://www.innocenceproject.org/causes/index.php (last accessed Aug. 26, 2003).

14. According to the Innocence Project, *supra* note 10, more than two-thirds of the 138 wrongful convictions were based, at least in part, on mistaken eyewitness identification. *Case Profiles*, The Innocence Project, *at* http://www.innocenceproject.org/case/index.php (last visited Aug. 26, 2003).

15. *See, e.g.*, Max M. Houck & Bruce Budowle, *Correlation of Microscopic and Mitochondrial DNA Hair Comparisons*, 47 J. FORENSIC SCI. 964 (2002), *available at* http://aafs.micronexx.com/PDF/JOFS/JFS2001398_475/JFS2001398_475.pdf (last accessed Nov. 22, 2003).

16. Note that nuclear DNA from hair, if it can be extracted and analyzed, will serve to individualize the hair's source. Even mitochondrial DNA will have a powerful limiting effect on who the hair donor could have been, and an absolute effect on eliminating specific individuals. Where hair is involved, the defining dichotomy is between DNA analysis and microscopic analysis.

17. 522 U.S. 136 (1997).

and *Kumho Tire Co. v. Carmichael*,[18] in order to present a non-orthodox, but perhaps a clearer way of regarding these cases. Section II will also discuss the manner in which the several states have reacted to *Daubert*. Many have followed it, but many have rejected it and retained *Frye*. A few states have rejected both, asserting that no reliability requirement should be imposed on expert testimony. This article argues that, in spite of references to reliability in *State v. Williams*, and contrary to lawyerly folk belief that Maine is a *Daubert* state, Maine in fact belongs to this last set of states, those that decided to "sit out the dance" or, depending on one's metaphorical preference, those marching to their own drummer(s). In several of these states, (and Maine appears to be one of them) a deliberate choice seems to have been made to maintain a liberal admission policy for scientific evidence, either by expressly avoiding a reliability requirement, or by stating the reliability requirement in such a manner that it is unlikely to be used to exclude evidence.

Section III will discuss specific types of forensic evidence, both from a scientific point of view and a case law point of view. There are many types of tests used in forensic investigations, including some that hang on in spite of having long been recognized as bogus by the scientific community, and others that were once considered valid but are now realized not to be. In addition, there are a number of forensic techniques, including fingerprint and handwriting identification, that have been used for years without having been tested for accuracy or reliability.[19] It is not clear which, if any, of these latter techniques will ultimately be found to fail the *Daubert* test for reliability. What *is* clear is that this last set of techniques has given rise to the greatest amount of *Daubert*-induced rancor. In some cases, such as with latent fingerprint identifications, the present practitioners have taken up a very defensive stance, either claiming that no validation tests need be done or that they have already been done decades ago. In contrast, practitioners of other techniques said to be unsupported by validation tests have been cooperative and open in developing both general validation tests and individual proficiency tests. The forensic techniques and devices selected for more detailed discussion in Section III are: (a) polygraph; (b) latent fingerprint identification; (c) "voiceprints"; (d) handwriting identification; and (e) bullet "fingerprinting" by trace element analysis.

Section IV will be a polemic, arguing that: (1) the *Daubert* standard regarding expert testimony reduces essentially to that of "general acceptance within the scientific community," one of the versions of the *Frye* standard; (2) determining whether the scientific community as a whole accepts a theory or technique is readily knowable and does not involve "counting heads"; (3) for the admission of testimony that is ostensibly scientific into evidence, there *should* be a reliability requirement and the requirement should be articulated in such a way that its application can be evaluated by an independent observer;[20] (4) Maine lacks such a stan-

18. 526 U.S. 137 (1999).

19. More carefully stated, it has been realized that these techniques have not been validation-tested in a manner that can be evaluated by disinterested observers who are able to evaluate the statistics and other goodness measures of the validation tests.

20. There must be enough structure to the reliability test that someone aggrieved by the admission or exclusion of scientific testimony at trial has something on which to structure his or her appellate argument. The present *Daubert* standard comes close to satisfying this requirement.

dard; and (5) in trials of criminal cases, the defense must be required to engage a forensic expert if the prosecution's case is based in any way on forensic testimony or on evidence that can be evaluated by a forensic scientist.

II. THE *FRYE* AND *DAUBERT* STANDARDS, AND THE STANDARDS OF MAINE AND OF THE OTHER FORTY-NINE STATES

A. Frye *Standard*[21]

Mr. Frye was convicted of second-degree murder in the federal district court for the District of Columbia after a trial in which his exculpatory lie detector[22] findings were not allowed into evidence.[23] This version of the lie detector just measured a single physiological parameter, systolic blood pressure, and the assertion of its inventor-operator was that the pattern of blood pressure variation allowed him to determine whether the subject was being truthful.

Ironically, in terms of later developments, the trial court's stated grounds for excluding the testimony was relevance; the test had been done ten days before trial and therefore was not relevant to the truthfulness of Mr. Frye's trial testimony.[24] It was at the appellate level that the grounds for exclusion were stated in terms of the quality of the evidence rather than its relevance. The D.C. Court of Appeals set out, as the common law rule pertaining to expert witnesses, the following (which is now familiar from Rule 702):

> The [common law] rule is that the opinions of experts or skilled witnesses are admissible in evidence in those cases in which the matter of inquiry is such that inexperienced persons are unlikely to prove capable of forming a correct judgment upon it, for the reason that the subject-matter so far partakes of a science, art, or trade as to require a previous habit or experience or study in it, in order to acquire a knowledge of it. When the question involved does not lie within the range of common experience or common knowledge, but requires special experience or special knowledge, then the opinions of witnesses skilled in that particular science, art, or trade to which the question relates are admissible in evidence.[25]

Following directly on that statement, the court said

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.[26]

---

21. It may be thought that a great deal of space is given to a standard that has been "superseded." The attention paid to *Frye* in this discussion is justified because of its continuing health as the scientific evidence standard in many states. Furthermore, many consider its incorporation into *Daubert* as the most important of the "*Daubert* factors." Indeed, this Author believes that the "*Daubert* factors" collectively reduce to the single "general acceptance" factor for the majority of the types of scientific evidence sought to be introduced.

22. In this Article, the use of the term lie detector should not imply that it is accepted as a descriptive label for polygraph.

23. *Frye v. United States*, 293 F. 1013, 1013-14 (App. D.C. 1923).

24. *Id.* at 1014. *See infra* Part III.A.

25. *Frye v. United States*, 293 F. 1013 at 1014.

26. *Id.*

Finally, the court applied this rule to the case before it, saying, "[w]e think the systolic blood pressure deception test has not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made."[27]

It is important to note, in light of later complaints that the *Frye* standard created too many uncertainties as to which branch of science was the "particular" one that trial judges should look to, that the *Frye* court had anticipated a broad determination. Critics pointed out that when a narrow definition was used, more evidence came in than when a broad definition was used.[28] It is clear that at the outset it was a broad consensus that was required. One can find an acceptance of physiology-based lie detectors among the group that uses them or advocates their use. Some of these people may have the credentials of scientists. However, there has not been general acceptance of this type of lie detector among either the fields of physiology or psychology, at least not within clinical psychology.

Also important to note in the *Frye* court's framing of the standard is the exact meaning of the phrases used. The court is not saying that just because a form of evidence has long been accepted it will continue to be accepted. The "thing" from which a deduction is made must be sufficiently established to have gained general acceptance in the scientific field in which it resides. It does not matter if the subject matter of the expert testimony has been repeatedly introduced into courts; it shall be excluded if it does not have general scientific acceptance. This allows for the conclusion that something, assumed to have general acceptance because it has been around for a long time, in fact does not have acceptance, nor merit it. It also allows the court to take notice that scientific acceptance has been withdrawn from a "thing" that once had it.

Over the following seventy years, the *Frye* standard in various forms was adopted by most federal and state jurisdictions in the United States.[29] Although in many cases, the adopting language implied that a narrow branch of science was to be looked to for the "general acceptance" measure,[30] other courts were clear about the breadth of the view. In some cases, it was the scientific community as a whole that was to be the base, making the review broader than even the *Frye* court had specified.[31] New York State, for example, continues to adhere to the version of *Frye* that it expressly adopted in 1983 in *People v. Hughes*.[32] A subsequent New York decision, in accord with *Hughes*, stated that, "[w]hile foundation concerns itself with the adequacy of the specific procedures used to generate the particular evidence [sought] to be admitted, the test pursuant to [*Frye*] poses the more elemental question of whether the accepted techniques, when properly performed,

---

27. *Id.*

28. *See, e.g.*, Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence:* Frye v. United States, *A Half-Century Later*, 80 COLUM. L. REV. 1197, 1208-09 (1980).

29. *See supra* note 4. *See also infra* note 72 and accompanying text.

30. *See, e.g.*, Commonwealth v. Fatalo, 191 N.E.2d 479, 481 (Mass. 1963) ("Judicial acceptance of a scientific theory or instrument can occur only when it follows a general acceptance by the community of scientists involved.").

31. Note in the discussion of the polygraph, *infra* Part III.A, that the recent National Research Council evaluation was done by a committee that included scientists outside of the fields of physiology and psychology, including in particular statisticians.

32. 453 N.E.2d 484, 497 (N.Y. 1983).

generate results accepted as reliable within the scientific community generally."[33]

The language in this ruling serves to answer those who assert that the major criticism of *Daubert* is internally inconsistent. The alleged inconsistency lies in claiming that most judges will be unable to recognize science when they see it, while at the same time contending that the *Frye* "general acceptance" rule should be the standard. How, they ask, can a judge, whose lack of scientific understanding is the supposed problem, choose the correct group of scientists to look to in applying the "general acceptance" test, or, for that matter how can such a judge determine whether the chosen group consist of scientists in the first place. The specter is raised of a judge looking for "general acceptance" within an insular and secretive group of practitioners who are the only ones working with the technique in question, for example the correlation of people's psyches with the moon phases, or of the discerning of truth from the examination of donkey entrails. Scientists outside the group would ridicule this "science," but since they are outside the group deemed relevant, the judge, under *Frye*, would not be asking them. Indeed, with borderline cases, the smaller the relevant group defined by the judge the more likely the testimony would be admitted, and conversely, the broader the judge has defined the group the less likely that it would be admitted.

The recognition that it is the *broad* scientific community that should be looked to eliminates the problem of which "community" to choose. In addition, the statement of the rule in terms of the techniques that were actually used cuts through the formalistic objection that continues to be raised by those who question the broad applicability of *Frye* to all scientific evidence. Critics point out that the *Frye* standard refers to a generally accepted "doctrine or theory" underlying the testimony, and ask where that leaves the technique that, though universally accepted, is not yet explainable by any theory, let alone one accepted by a broad consensus of scientists. However, it is clear from reading *Frye* that the "thing" from which the deduction is made is not limited to a deep underlying principle, but can include the technique itself.[34]

Another criticism leveled at the *Frye* standard is that it requires judges to look outside the law in order to determine whether expert testimony is admissible. In the wake of his *Structure of Scientific Revolutions*,[35] Thomas Kuhn received letters from social scientists, historians and others not in the hard sciences informing him that "paradigms" existed in their fields also. He noted that it was difficult to respond to people who missed the point by such a wide margin.[36] Unless the

33. People v. Wesley, 633 N.E.2d 451, 454 (N.Y. 1994).

34. For the decades before Bardeen, Cooper, and Shreiffer provided the now-universally accepted BCS theory, the most common form of superconductivity (the phenomenon by which an electrical conductor loses absolutely *all* resistance to current flowing through it), no one could explain superconductivity. In other words, there was no "theory or doctrine" underlying it. If the *Frye* standard required a generally accepted "theory or doctrine," it would appear that testimony based on superconductivity would not have been allowed into evidence. Of course, in reality, that paradox would be easily resolved by a trial judge governed by the original *Frye* standard. The evidence would be admitted, and a slight adjustment would be made in the application of the rule. In New York, that problem is resolved globally by making the technique itself the element that has to have general scientific acceptance. *See supra* note 33 and accompanying text.

35. THOMAS KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS (1962).

36. THOMAS KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS 208 (2d. Ed. 1969) ("I see what they mean . . . but their reaction has nevertheless puzzled me.").

judge is willing to ignore the reliability of material, he or she must look to the field from which it comes. One of the more unenlightened forms of *Frye* bashing is to depict it as simply calling for a counting of heads within a subset of the scientific community.[37] Unlike a jury verdict, scientific consensus is not arrived at by a vote. There must be reasons amenable to logical analysis set out when one states why a theory or a result is wrong or doubtful. The reports generated by the National Research Council give good examples of how the scientific community goes about evaluating a theory, and these reports are written in language accessible to everyone.[38]

In articulating their version of the *Frye* standard, some state courts recognized that a determination of the reliability of scientific results requires three levels of inquiry.[39] For present purposes, these three levels can be labeled, respectively, *Frye I, Frye II,* and *Frye III.* At the most fundamental level is the underlying scientific doctrine; the *Frye I* question is whether that doctrine has general scientific acceptance. As a practical matter, this is usually the least important inquiry, though often the one given the most attention. Rarely, if ever, does the underlying scientific doctrine lead directly to the scientific results that are sought to be introduced into evidence. For example, every analysis dealing with the motion of objects and their interaction with one another is based on Newton's Theory of Motion, also referred to as Newton's Laws. However, it is specious to argue that expert testimony should be admitted just because the witness asserts that it was based on Newton's Laws. Thus, it is necessary to go to the *Frye II* level to determine whether the technique has general scientific acceptance.

A common but simple technique in accident reconstruction is to determine the speed of a car at the start of its skid by measuring the distance it took to stop, known as the skid-to-a-stop distance. It can be shown from Newton's Laws that in order to calculate the speed, it is only necessary to measure the length of the skid (DIST) and determine the coefficient of friction (COF) between the tires and the pavement. Furthermore, there is general scientific acceptance of the fact that for a given dry pavement surface the COF is the same for all over-the-road automobile tires in use in the United States.[40] Because of this acceptance, one can measure COF with any car and assume that the result will apply to the car in question. One expression for the needed relationship is:

$$V = [30 \times DIST \times COF]^{1/2}$$

V is the speed of the car in mph at the start of the skid if DIST is stated in feet.[41] It can be shown from Newton's Equations that the same expression holds

37. *See, e.g.*, Harper v. State, 292 S.E.2d 389 (Ga. 1982).

38. See *infra* notes 92 and 187 for the National Research Council reports on polygraphs and voiceprints, respectively.

39. *See, e.g.*, Ex parte Perry, 586 So.2d 242 (Ala. 1991); State v. Schwartz, 447 N.W.2d 422 (Minn. 1989); House v. State, 445 So.2d 815 (Miss. 1984). Requiring two of the three levels of inquiry were State v. Stout, 478 S.W.2d 368 (Mo. 1972) and State v. Coolidge, 260 A.2d 547 (N.H. 1969).

40. The range of tires on standard vehicles is such that the tire/road COF is the same for all such cars on a given pavement, something that is not intuitively obvious.

41. If the expression giving the speed-from-skids is stated in a different unit system, it will have a different appearance in terms of superficialities. However, if the expression is correctly derived, it can be easily shown to be the same, physically, as the one given.

for all cars, regardless of weight, suspension, or other specific design features. The determination of speed using this equation and the assumption that the COF is essentially the same for all car tires constitutes this particular technique, and is universally accepted both among persons carrying out these measurements and by objective scientists evaluating the technique. Thus, this technique satisfies *Frye II*, and in fact is used in court hundreds of times a year.

The *Frye III* level of inquiry goes to how well the particular practitioner applied the technique. In the example above, this would relate to how well he or she measured the skid distance and the COF. Regardless of arguments about the competence of the practitioner involved, one does not get into whether there would be general acceptance of the particular application. That is considered when weighing the evidence and is generally dealt with on cross examination.

There is some confusion about what constitutes a novel technique. To continue with the example given above, if the pavement surface is wet, to the point where there is a film of standing water, it is no longer true that the COF will be the same for all tires. It will be much lower for bald tires, for example, than it will be for tires containing the legally required 1/16" tread. There is not, therefore, general acceptance of the technique as described above when there is water on the pavement, and it would be an error for a trial court to accept the profferer's statement that it was. If the court did accept the technique and equation at face value, even though the road was wet at the time of the event under consideration, the speed calculated for the skidding car could be significantly higher than its actual speed if that car had had tires with little tread.

For another example from the arena of accident reconstruction, consider "speed from yaw."[42] This is a good example since it illustrates a technique that is accepted (given that certain conditions are satisfied), yet the theory underlying it is not completely understood. It can be derived from Newton's Equations that the *maximum* speed with which a car can round a curve of radius R (that is, an arc of a circle with that radius) is:

$$V_{MAX} = 3.86x[RxCOF]^{1/2}$$

V_{MAX} is given in mph if R is given in feet. COF is still a measure of the resistance needed to slide the car's tires on the pavement and, as such, is the same for all cars on the particular pavement traversed by the car, again provided that the pavement is dry. By saying that V_{MAX} is the maximum speed is to say that attempts to travel the trajectory at higher speed will be foiled; it simply cannot be done. Now, to persons first encountering this expression, the most amazing thing is that it does not depend at all on the make of the car or the weight of the car. This was also true of the skid-to-a-stop relationship. However, it does not seem reasonable that the maximum speed at which a car can round a curved path does not depend on such factors as the suspension, the stiffness of the springs and so forth. But it *is* true; the above expression is rooted very close to Newton's Equations, with no intermediate steps. What is not yet understood, however, is why the speed at which a car starts to sideslip when going around a curve of radius R is approximately *equal* to V_{MAX}. Nothing in the theory as developed thus far explains why

42. *See* J. STANNARD BAKER & LYNN B. FRICKE, THE TRAFFIC-ACCIDENT INVESTIGATION MANUAL: AT-SCENE INVESTIGATION AND TECHNICAL FOLLOW-UP 17-19 – 17-22 (1986).

that is.[43] Nevertheless, experimental tests carried out by both police and civilian agencies have led to the general scientific acceptance of this technique for determining the speed that a car must have had when it was leaving cornering marks on the pavement with a radius of curvature R. The key equation in which V_{MAX} gives the "critical speed" for a car rounding a curve of radius R when the COF has the value stated is sometimes referred to as the "yaw formula" or the "critical speed scuff formula."

On the other hand, when the yaw formula is used in an attempt to calculate speed for a car that left a trace of its passage on something other than a hard surface, the formula (that is, the technique) is *not* being used in a context that has been generally accepted by the scientific community and, under *Frye*, the results should be excluded from evidence. However, Maine has allowed the admission of such evidence.[44]

B. Daubert *Standard*

In the early 1990s, amidst expressions of dismay that nonsense was being accepted as science in the courtrooms of the nation[45] and equally adamant protestations that the *Frye* standard was so strict that it was unfairly keeping out perfectly valid scientific results, several cases—ultimately combined under the caption *Daubert v. Merrell Dow*[46]—were making their way to the U.S. Supreme Court on a scientific evidence issue. Because this was the first time that the Supreme Court was to address scientific evidence admissibility, many diverse interests, cutting across political boundaries, submitted *amici curiae* briefs.[47] Although many of those submitting briefs are readily recognizable, others are not, and were probably ad hoc groups brought into existence to exercise an opportunity to be heard by the Court. Briefs supporting the Petitioners in urging that "strict" *Frye* standard be struck down and that it be made easier to introduce novel scientific findings into evidence were submitted by, among others, (a) The American Society of Law, Medicine and Ethics, (b) the Association of Trial Lawyers of America (ATLA), (c) Nicolaas Bloembergen, et al.,[48] and (d) a collective of four states—Texas, Montana, Idaho, and South Dakota. Briefs arguing either for the retention of the *Frye* standard or for a stricter standard of admissibility were submitted by, among others, (a) the American Insurance Association (AIA), (b) the Pharmaceutical Manufacturers Association (PMA), (c) the American Association for the Advancement of Science (AAAS) joined with the National Academy of Sciences (NAS), (d) The Chamber of Commerce of the United States, (e) The American Tort Reform Association, (f) the United States government, submitted by the Office of

43. *See, e.g.*, 1 FORENSIC ACCIDENT INVESTIGATION 52 (T.L. Bohan & A.C. Damask, eds., LexisLaw, Supp. 2003) (making the somewhat hyperbolic analogy to the surprise one would have in discovering that the energy emitted when an object of mass M burned in oxygen is equal to the maximum energy theoretically available from the object, as given by the famous E = Mc2.

44. *See* State v. Irving, 2002 ME 31, 818 A.2d 204.

45. The catch phrase was "junk science," probably popularized by PETER W. HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991), largely a propaganda tract biased toward large defendants, the best part of which is its title.

46. 509 U.S. 579 (1993).

47. *Id.* at 581.

48. This group characterized themselves as eighteen scientists, scholars, and teachers of science.

Solicitor General, (g) a consortium that may have included every medical association in the United States, and (h) a group of scientists headed by Nicolaas Bloembergen, most of whom were Nobel Prize laureates.

Finally, some organizations submitted "neutral" briefs including: (a) the Carnegie Commission on Science, Technology, and Government and (b) "A Group of American Law Professors."[49]

The trial court granted summary judgment for Merrell Dow after using the *Frye* standard to exclude the Plaintiff's key expert testimony.[50] The United States Supreme Court vacated summary judgment.[51] The Court's opinion, prepared by Justice Blackmun, was to some observers, especially those in the sciences, a call for all judges to become scientists themselves, or at least to act as if they were scientists in evaluating proffered scientific testimony. There was no longer an absolute dependence on what the scientific community, however it was defined, thought of the science that was being offered in court. Instead, courts were to determine whether the testimony was based on the scientific method, placing their emphasis on the method by which the results were said to have been obtained, and minimizing the importance of those results themselves as a measure of reliability.[52]

In summary, *Daubert* emphasized the primacy of Rule 702 in determining whether expert scientific testimony was to be admitted, but also imposed a requirement that the Court had inferred from Rule 702, namely, that the testimony be reliable.[53] As guidance to the trial judges on whom it was placing the responsibility of gauging reliability, the Court listed some factors (including "general acceptance") that might be used to evaluate the evidence, while cautioning that no one of these factors nor any particular combination of them was dispositive for the reliability determination.[54] The implication was that the trial judge, after applying various measures, would know scientific reliability when he or she saw it. There are a number of ways of stating the *Daubert* factors. The following is a common one, where the "technique" refers to the means by which the results sought to be testified to were obtained:

---

49. In the Author's opinion, the AAAS/NAS brief urging the requirement of clearly articulated high admission standards was far above the other briefs in its clarity and in the statement of its arguments. Apparently, the Supreme Court was also impressed, since much of the text in the opinion paraphrased or borrowed verbatim from this brief. *See* Brief of Amici Curiae American Association for the Advancement of Science and National Academy of Sciences, Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993) (No. 92-102). However, it can be argued that the *substance* of the opinion fell far short of what the AAAS and NAS sought from the Court.

50. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. at 583-84.

51. *Id.* at 597-98.

52. As an aside, it should be said that this is not the way that scientists themselves decide whether a new result is legitimate. Scientists, especially experimental scientists, realize that mistakes can be made easily and that an individual experimentalist may be mistaken in his or her belief about how the experiment was done. For this reason, it is the *result* as well as the stated experimental method that is examined. If the result is sufficiently weird or otherwise unexpected, considerable time will go by before it is accepted, time during which others try to replicate the experimental results. It is likely that Justice Blackmun was influenced by a statement in the *amicus* brief filed by the Nobel laureates, which contains the observation, "[i]t is how the conclusions are reached, not what the conclusions are, that makes them 'good science' . . . ." One wants to respond, "Yeah, but. . ."

53. Daubert v. Merrell Dow Pharm., Inc., 504 U.S. at 594-95.

54. *Id.* at 592-94.

(a) Is the technique testable and tested?

(b) Has there been publication and peer review of the technique?

(c) Is the rate of error obtained by the technique reasonable?

(d) Are there standards for applying the technique?

(e) Is there general scientific acceptance of the technique?[55]

Rule 702 applies to all expert opinion testimony and not just to scientific testimony. The emphasis of *Daubert* on *scientific* testimony, however, led to wishful thinking by many litigants that the strictures of *Daubert* did not apply to the type of opinion testimony that *they* were going to rely on. Indeed, some appellate courts supported this view,[56] leading to such unseemly behavior as attorneys cautioning their expert witnesses not to call what they did "science."[57] It was not just attorneys, but also forensic scientists who were assured that *Daubert* would never apply to most of them because it was only concerned with "breakthrough" discoveries.[58] Science was one thing. Experience was another. It was argued that *Daubert* applied only to those trying to introduce evidence based on new scientific discoveries and techniques, not to scientists testifying about results produced by "normal science,"[59] and certainly not to engineers.[60] It appears that no one even speculated that *Daubert* would apply to the opinions of physicians providing expert testimony based on their medical experience, or to accountants, bankers, and mapmakers giving expert testimony pursuant to Rule 702.

Any hope that *Daubert* could be narrowly applied was dashed by the Supreme Court in *Kumho Tire Co. v. Carmichael*.[61] Like *Daubert*, *Kumho Tire* was watched very carefully by those interested in expert testimony. The issue before the Court was whether Rule 702 guidelines and exhortations of *Daubert* were limited to scientific testimony, or whether they also applied to engineering testimony, particularly engineering testimony based solely on the professional experience of the expert witness.[62] *Amicus curiae* briefs were filed, including a major one on behalf of engineers arguing that engineering was different from science and therefore should not be held to the formal reliability standards of *Daubert*.[63] The opinion in *Kumho Tire*, though consistent with the breadth of Rule 702, was surprisingly broad, holding that *all* expert opinion testimony governed by Rule 702 was also governed

55. *Id.* at 592-95.

56. *See, e.g.*, Carmichael v. Samyang Tire, Inc., 131 F.3d 1433 (11th Cir. 1997), *rev'd* Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

57. The National Fire Protection Association (NFPA) received more than fifty proposals immediately following *Daubert* asking that references to the "scientific method" be removed from NFPA 921 *Guide for Fire and Explosion Investigation*. Private communication with John Lentini, Principal Member of NFPA Technical Committee on Fire Investigation (July 2003).

58. At the Plenary Session of the 1995 Annual Meeting of the American Academy of Forensic Sciences (AAFS), former AAFS President Cyril Wecht forcefully delivered the message that no one in that room would ever have to confront a *Daubert*-based hearing. Cyril Wecht, M.D., J.D., Comments at the American Academy of Forensic Sciences Annual Meeting (Feb. 15, 1995).

59. *See* generally KUHN, *supra* note 35.

60. *See* Bogosian v. Mercedes-Benz of N. Am., Inc., 104 F.3d 472 (1st Cir. 1997) (holding that a forensic automobile mechanic's testimony was not scientific and that *Daubert* did not apply).

61. 526 U.S. 137 (1999).

62. *Id.* at 141-42.

63. Brief of Amicus Curiae National Association of Forensic Engineers, Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999) (No. 97-1709).

by the reliability demands of *Daubert*.[64] As a result, *Daubert* applies to *all* expert opinion testimony, regardless of its particular specialized nature.[65] While suggesting that the specific factors provided in *Daubert* would not be as applicable to the reliability evaluations of many non-scientific types of expert testimony, the *Kumho Tire* Court nevertheless made it clear that some form of reliability evaluation of all proffered expert testimony had to be conducted by the trial judge.[66]

Rounding out the *Daubert* trilogy is *General Elec. Co. v. Joiner*,[67] in which the United States Supreme Court established that in federal jurisdictions, the standard of appellate review of a trial court decision regarding the admission or exclusion of expert testimony (under *Daubert*) would be abuse of discretion.[68] For appellants seeking to reverse a decision made by the trial court, abuse of discretion is the most difficult appellate standard of review to confront.

C. The Standard in Maine and the Other States

At the time that *Daubert* was granted *certiorari* by the United States Supreme Court, most states and federal circuits that articulated a scientific evidence standard used some form of *Frye*, but Maine did not. The governing case in Maine, *State v. Williams*,[69] is similar in important ways to *Daubert*, which it preceded by fifteen years, but with one crucial difference. In explicitly disavowing the "general acceptance" rule and directing trial judges to look exclusively to Maine Rule of Evidence 702 (Maine Rule 702),[70] neither *Williams* nor any subsequent Maine Supreme Judicial Court decision has provided practical guidance for the application of that rule to proffered scientific evidence. This seems not to be an oversight, but rather to reflect a leaning in Maine toward admitting evidence rather than excluding it.

A recent survey[71] asserts that as of July 2001 twenty-five states adhered to

64. Kumho Tire Co. v. Carmichael, 526 U.S. at 141.

65. *Id.*

66. Many, if not most, states have rules of evidence similar or identical to Rule 702. For example, Maine Rule 702 is verbatim to the federal version of 702 in its pre-*Daubert* form. M.R. EVID. 702.

67. 522 U.S. 136 (1997).

68. *Id.* at 142-43.

69. 388 A.2d 500 (Me. 1978). Maine cases that follow *Williams* include: State v. Irving, 2003 ME 31, 818 A.2d 204; In re Jon N, 2000 ME 123, 754 A.2d 346; State v. Kelly, 2000 ME 107, 752, A.2d 188; State v. Tomah, 1999 ME 109, 736 A.2d 1047; State v. Poulliot, 1999 ME 39, 726, A.2d 210; State v. MacDonald, 1998 ME 212, 718 A.2d 195; State v. Fleming, 1997 ME 158, 698 A.2d 503; State v. Taylor, 694 A.2d 907 (Me. 1996); State v. Stade, 683 A.2d 164 (Me. 1996); Kay v. Hanover Ins. Co., 677 A.2d 556 (Me. 1996); State v. Harper, 675 A.2d 495 (Me. 1996); State v. Ellis, 669 A.2d 752 (Me. 1996); State v. Grigsby, 666 A.2d 503 (Me. 1995); State v. Kennedy, 657 A.2d 773 (Me. 1995); State v. Chapman, 645 A.2d 1 (Me. 1994); State v. Boobar, 637 A.2d 1162 (Me. 1994); State v. Roman, 622 A.2d 96 (Me. 1993); State v. Boutilier, 426 A.2d 876 (Me. 1981).

70. M.R. EVID. 702. As indicated above, Maine Rule 702, exactly the same as the corresponding federal rule at that time, has not been amended to bring it into compliance with the federal Rule 702 as amended to reflect *Daubert*.

71. Alice B. Lustre, *Post-Daubert Standards for Admissibility of Scientific And Other Expert Evidence In State Courts*, 90 A.L.R.5th 453, 454 (2001).

Daubert, sixteen were still in the *Frye* camp,[72] six used a combination of, or alternated between the two depending on the nature of the evidence, and four followed rules of their own devising, and do not fit into any of the previous three categories.[73]

The four states that the survey identifies as having devised non-*Frye* non-*Daubert* rules are Wisconsin, Virginia, Georgia, and Utah.[74] The first three explicitly exclude the *Daubert* emphasis on reliability; Utah is similar to *Daubert* in requiring a reliability check, but seems to apply it more stringently, with the effect that it tends to exclude evidence that would probably come in under *Daubert*.[75] I would also place Maine in the non-*Frye*, non-*Daubert* category.

The governing Wisconsin case, *State v. Walstad*,[76] is pre-*Daubert* and expressly rejects *Frye*, stating, "[n]owhere in the Wisconsin Rules of Evidence[77] or in the extensive commentaries to it is the *Frye* rule mentioned. Under our rules, if the evidence is relevant, it is admissible, unless it is excluded for some special reason, such as prejudicial effect or jury confusion."[78] Unlike *Daubert* and its progeny, *Walstad* does not find a "reliability" condition lurking within Rule 702. Without that, Rule 702 is seen to be a *permissive* rule, rather than a restrictive one. It permits expert opinions to be admitted as evidence; apart from the minimal condition of relevance, it places no restrictions on such admission. Consistent with the *Walstad* rejection of *Frye* because of its precondition of reliability, an intermediate Wisconsin appellate court subsequently rejected *Daubert* explicitly.[79] Indicating that the gate is wide open, a recent Wisconsin Supreme Court case observed in a footnote that:

---

72. With respect to trends, a paper this Author co-authored early in the post-*Daubert* era listed the late-1994 breakdown to be twenty-seven states following *Frye*, nine following *Daubert*, and the rest not well defined. Thomas Bohan & Erik Heels, *The Case Against Daubert: The New Scientific Evidence "Standard" and the Standards of the Several States*, 40 J. FORENSIC SCI. 1030, 1035-36 (1995). Five states, including Maine, were identified at that time as either having failed to set out a clear standard or as nominally following one and actually following the other. *Id.* at 1036. Thirty-eight states had by that time adopted a formal rule of evidence that tracked federal Rule 702 in the form it had before the *Daubert*-inspired amendment. *Id.*

73. It is interesting that Lustre places Maine in the *Daubert* column in spite of the Maine Supreme Judicial Court having indicated repeatedly that it is not. Lustre, *supra* note 71, at 502. *See also* RICHARD H. FIELD & PETER L. MURRAY, MAINE EVIDENCE § 702.3 (3d ed. 1994), which insists that *State v. Williams* governs Maine and that *State v. Williams* is not *Daubert*.

74. Lustre, *supra* note 71, at 543-45.

75. *Id.* These conclusions are based on State v. Rimmasch, 775 P.2d 388 (Utah 1989), wherein the Utah Supreme Court, citing with approval its earlier commentary in Phillips v. Jackson, 615 P.2d 1228, 1234 (Utah 1980), ruled that Utah's adoption of Rule 702 did not do away with "threshold reliability requirements," and stated "'[a]n analysis of the admissibility of scientific evidence, while taking into account general scientific acceptance and widespread practical application, must focus in all events on proof of inherent reliability.'" State v. Rimmasch, 775 P.2d at 397 (quoting Phillips v. Jackson, 615 P.2d at 1234).
However, in 2002, the Utah court clarified its *Rimmasch* decision, stating,"we reaffirm our previous holdings that the *Rimmasch* test applies only to novel scientific methods and techniques. Other scientific testimony is to be evaluated under rule 702 without heightened tests of 'inherent reliability.'" Alder v. Bayer Corp., 61 P.3d 1068, 1084 (Utah 2003).

76. 351 N.W.2d 469 (Wis. 1984).

77. The Wisconsin Rules of Evidence track the Federal Rules of Evidence.

78. State v. Walstad, 351 N.W.2d at 485-86.

79. State v. Peters, 534 N.W.2d 867 (Wis. Ct. App. 1995).

> In the federal system, trial courts have a significant "gatekeeper" function in keeping from the jury expert testimony that is deemed unreliable. By contrast, the trial court's gatekeeper role in Wisconsin is limited. In Wisconsin, "[o]nce the relevancy of the evidence is established and the witness is qualified as an expert, the reliability of the evidence is a weight and credibility issue for the fact finder and any reliability challenges must be made through cross-examination or by other means of impeachment."[80]

Virginia is another laissez faire state regarding any reliability test of scientific evidence, and apparently has never reversed a judgment based on the improper admission of scientific evidence. It expressly rejected *Frye* in 1988, in *O'Dell v. Commonwealth*.[81] Although in 1990, the Virginia court, while reaffirming its rejection of *Frye*, acknowledged that a threshold evaluation of reliability was desirable, it did nothing to implement such evaluations, and simply directed trial judges to determine reliability through expert testimony, with no guidelines as to what to look for from that testimony.[82] Nevertheless, the Virginia high court has not ruled out the possibility of its adopting *Daubert* at some point. In *John v. Im*,[83] the Virginia court stated:

> [W]e do not reach the merits of the issue whether that evidence also failed to meet the criteria for scientific reliability articulated in *Daubert*. We note, however, that we have not previously considered the question whether the *Daubert* analysis employed by the federal courts should be applied in our trial courts to determine the scientific reliability of expert testimony. Therefore, we leave this question open for future consideration.[84]

Finally, Georgia has a standard that is effectively similar to Maine's *de facto* standard. In *Harper v. State*,[85] it rejected *Frye* as tantamount to "counting heads," and concluded:

> The *Frye* rule of "counting heads" in the scientific community is not an appropriate way to determine the admissibility of a scientific procedure in evidence. . . . [W]e hold that it is proper for the trial judge to decide whether the procedure or technique in question has reached a scientific stage of verifiable certainty, or in the words of Professor Irving Younger, whether the procedure "rests upon the laws of nature."[86]

This so far remains good law.[87] However, the Supreme Court of Georgia recently granted *certiorari* to hear *Orkin Exterminating Corp. v. Carder*,[88] to consider "What

---

80. Conley Publ'g Group Ltd. v. Journal Communications, 665 N.W.2d 879, 892-93 n.21 (2003) (citations omitted) (quoting State v. Peters, 534 N.W.2d 867 (Wis. Ct. App. 1995); citing Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993), Kumho Tire Co. v. Carmichael, 562 U.S. 137 (1999)).

81. 364 S.E.2d 491 (Va. 1988).

82. Spencer v. Commonwealth, 393 S.E.2d 609 (Va. 1990).

83. 559 S.E.2d 694 (Va. 2002).

84. *Id.* at 697-98. The court also noted that "[p]rior to *Daubert*, however, we discussed the trial court's role in making a threshold finding of scientific reliability when unfamiliar scientific evidence is offered." *Id.* at 698 n.3 (citations omitted).

85. 292 S.E.2d 389 (Ga. 1982).

86. *Id.* at 395 (footnotes omitted).

87. *See* Pullin v. State, 534 S.E.2d 69, 70 (2000) ("In determining whether a scientific principle or technique has reached a stage of verifiable certainty that it is competent evidence in a court of law, *Harper v. State* . . . controls.").

88. 575 S.E.2d 664 (Ga. 2002), *cert. granted* No. 503G0650, 2003 Ga. Lexis 379, at *1 (Ga. Apr. 29, 2003).

standards or factors should govern the admissibility of expert scientific evidence in Georgia? Compare *Harper v. State* with *Daubert v. Merrell Dow Pharmaceuticals*, and *Kumho Tire Co. v. Carmichael*."[89] The governing statement regarding the Georgia standard is remarkable. It reflects an all too common misunderstanding about science and its practitioners, and overestimates the ability of persons not trained in technical fields to determine, with no input from scientists, whether a particular technique "rests upon the laws of nature." Although some of the language of *Daubert* reflects this attitude, it nevertheless provides guidance for evaluating most kinds of scientific evidence. Without that guidance, most questions of scientific reliability get left to the jury to decide, a jury that after viewing a "battle of the experts" is most likely to conclude that scientific opinion on the issue is evenly divided.[90] That is the situation, both in those states—such as Wisconsin and Georgia—that eschew a reliability requirement, and in those—such as Maine and Virginia—where no guidance is given in applying the requirement.

III. SPECIFIC TYPES OF FORENSIC EVIDENCE

In this Section, five different examples of forensic evidence will be discussed. The underlying science will be discussed for each type; as will be seen, each has a different status and degree of acceptance within the scientific community and with the courts.

A. Polygraph

1. The Science[91]

The frustration felt by the authors of the recent National Research Council polygraph report[92] sounds in their characterization of the device and its context:

> [S]everal themes in the polygraph debate have long histories: criticism by scientists of the scientific basis of polygraph testing, the development in the popular culture of a mystique of infallibility for polygraph lie detection, the use of the polygraph for security screening despite scientific criticism, policy debates lead-

89. *Id.*

90. One of the embarrassments of our judicial system is the provision it makes for lay juries to evaluate scientific claims out of context and to thereby arrive at conclusions that are diametrically opposed to those of the scientific community. This can easily happen when the jury is presented with two expert witnesses who disagree about the validity of a particular scientific point, not knowing that the entire scientific community is arrayed against one of the witnesses and on the side of the other. The whole point of judges acting as gatekeepers is to avoid putting juries in that position. To the extent that judges are unable to exercise this role when it comes to scientific evidence, justice is compromised.

91. Misleading as the lie detector label is, it provides more specificity than "polygraph," a completely general name referring to the instrument that measures several parameters at once and plots them with multiple pens on recorder paper. In contrast, the instrument at issue in *Frye* measured just one variable—systolic blood pressure—as a truthfulness gauge.

92. NAT'L RESEARCH COUNCIL, NAT'L ACADEMIES, THE POLYGRAPH AND LIE DETECTION (2003), *available at* http://www.nap.edu/books/0309084369/html (last visited Oct. 5, 2003) [hereinafter POLYGRAPH AND LIE DETECTION].

ing to decisions to end polygraph security screening programs, and debates over openness in polygraph research.[93]

Of all techniques that may be classified as scientific evidence, the lie detector appears to hold the record for the number of years over which it has been examined, as suggested by the above passage.[94] It would appear from case law that these studies have not been needed to educate the judicial system, but rather are requested because of the misplaced confidence that the public and governmental agencies place in them. It is as if each generation of public officials discover for the first time that lie detectors are not reliable, and in fact that there is no scientific evidence that they are able to detect lies about specific events or to serve as screeners of security risks or dishonest employees. And so, a new study is requested. The latest one was requested by the Department of Energy (DOE) in connection with pressure to carry out widescale security screenings of its employees.[95]

Unlike the public, the courts, beginning with *Frye*, have viewed lie detector tests with appropriate skepticism. As noted elsewhere, the *Frye* standard itself stems from a rejection of lie detector evidence in 1923.[96] Here it is noted that the enduring presence of polygraphs in the face of scientific and judicial rejection has two bases. One is the use that investigating agencies make of the polygraph to induce a suspect to confess, a use that is independent of whether the polygraph can actually detect lies. One aspect of its use in interrogation is, ironically, to lie about its results, playing on the suspect's faith in the instrument. This technique was used by the FBI in an interview with the Los Alamos physicist Wen Ho Lee, who was suspected of spying for China.[97] A major focus in the lengthy interrogation

93. *Id.* at 291. It is not possible to use the word "science" without thinking about the disputes in recent years regarding what is science and what is not. In spite of some commentators' belief that *Daubert* reduced the word to phenomena and techniques based on Newton's Laws of Motion, this discussion will follow the longer judicial tradition of referring to all practices that use specialized technical methods as being based on a science.

94. *Id.* at 1-6 (discussing "The Lie Detector Mystique").

95. POLYGRAPH AND LIE DETECTION, *supra* note 92.

96. *See supra* Part II.A.

97. Interview by FBI with Wen Ho Lee, suspected spy (Mar. 7, 1999), *at* www.wenholee.org (last visited Oct. 27, 2003) (accessed by clicking on the Documents item in the column on the left, and then clicking on "Declassified Transcript of FBI interrogation" under "FBI Documents"). The "identifier" (xxx) was used in the released transcript to indicate that one of the two FBI agents was talking. A typical portion of the transcript follows:

> (xxx): . . . do you know what's in this package? Do you know what's in the package that I got today and the phone call that I got from Washington? You failed your polygraphs.
> Lee: Okay.
> (xxx): You failed the one that you, you had in December. Okay? You failed that. The polygraph that (xxx) gave you?
> Lee: Uh-mm.
> (xxx): You couldn't pass it. When they asked you questions and they got down to issues about code issues, and they got down to weapons questions. You couldn't pass your polygraph.
> Lee: That, that, Washington, D.C., polygraph. They did not ask me anything about codes.
> (xxx): The Department of Energy polygraph?
> Lee: They did.
> (xxx): Whatever they asked you, you failed.

was the repeated lie that Dr. Lee had "failed" both polygraph tests administered to him earlier. Although this approach did not seem to work with Dr. Lee, it is indisputable that police lies during interrogation can induce innocent suspects to confess to crimes, as a recent example in Maine illustrates.[98] The polygraph then just becomes one more, albeit important, psychological weapon to be applied during police interrogations.

Unfortunately, the enduring use of the polygraph is also due to the fact that scientific evaluations of it over the years have been carried out primarily by advocates of the polygraph's use, and meant to demonstrate its efficacy. Because of its limitation, no tests yet performed have demonstrated that the polygraph detects lies. It would be far better if objective and scientifically acceptable tests of the instrument were carried out, ones that would determine either that the polygraph can detect lies (and if so, under what circumstances and with what degree of confidence) or, in the alternative, that it is no better than chance at detecting lies. Such tests are difficult to carry out, but not impossible, as is discussed in the recent National Research Council report. They would be very expensive, however, because of the noted difficulty, and is the type of undertaking that would only be supported by government funding. It would seem an appropriate use of funds by the U.S. Department of Justice (DOJ), and in particular by the DOJ National Institute of Justice (NIJ). A cynical explanation for why this has not been done is that powerful lobbies do not want to risk the likely outcome, which would remove the polygraph from the armaments of the police and security agencies. As will be discussed, the National Research Council found that, although there is no evidence that the polygraph can detect lies, there is evidence that when used in connection with interrogation concerning specific events, it is possible that the interrogation

---

Lee: . . . they only asked one question. Do you know what question is that?
(xxx): What was the question, WEN HO?
Lee: They only asked me. It's hard for me to say it, but it's something like uh, (pause). Did I pass information to somebody who can use that information to overthrow the U.S. Government (chuckles). Something like that. Do you understand?
(xxx): Well you failed! You failed that question!
Lee: How do you know I fail?
(xxx): I got it right here! The reports that they sent me! The video tapes when you were video taped in your polygraph? . . .
Lee: Yeah.
(xxx): . . . all the analysis that's been done in Washington! You failed WEN HO!
Lee: Why, why they told me . . .
(xxx): There's there's a black line and there's a white line. You either passed or you failed. You failed. When (xxx) talked to you in February? He asked you questions . . .
Lee: (Sighs).
(xxx): . . . And then (xxx) gave you a polygraph, you know what the results say in the second package? You failed that too. You failed everything!

Interview by FBI with Wen Ho Lee, suspected spy (Mar. 7, 1999), *at* http://www.wenholee.org.

98. *A False Confession to Murder*, at http://www.mtcforensics.com/confession.html (last visited Oct. 27, 2003). Raymond Wood was held in a York County jail for a year, awaiting trial, based on his "confession" during a lengthy interrogation in which he was told that the victim's hair had been found adhering to the underside of his van. These statements were not justified by the evidence at hand and in fact turned out to be false. When the judge viewed the interrogation tape, he threw out the confession and the State dropped the case.

has a rate of success in detecting untruths that is greater than what would be expected by chance.

In light of the statement that there has not been a scientific testing of the polygraph, one may ask how its effectiveness can be dismissed. In fact, all that can be said is that the polygraph has not been proven to detect lies or to confirm that the examinee is telling the truth. The scientific method, which courts have repeatedly held to be a necessary substrate for all scientific evidence, presumes that until something has been *shown* to be true, it is not accepted as true. In spite of the obvious value of this approach to all fields of knowledge, it appears that in many non-scientific fields and quasi-scientific fields, the opposite approach is adopted, where a statement is accepted as true as long as it has not been proven false.[99] Indeed, those who claimed that the *Frye* standard of general scientific acceptance placed too tight a requirement on the admissibility of scientific evidence should be comfortable with this philosophy. Such people fretted that the *Frye* test would prevent a modern-day Galileo from testifying based on his theories, because they would not have "general acceptance" within the scientific community. Even to the extent that the premise underlying the Galileo analogy is correct (and it probably is not),[100] a rule that permits any opinion witness to testify to his or her own idiosyncratic theory, its merit to be decided by lay jurors before it had been tested by the scientific community is what would make our system of justice fall into disrepute if not make it a laughing stock.[101]

The just-published report of the National Research Council regarding its evaluation of the polygraph is an excellent example of what a technique or theory must be to have the general acceptance of the scientific community. This work by the National Research Council is an indication of how the judicial system may connect with the scientific community in evaluating a wide range of scientific techniques and thus determine whether the respective techniques have general acceptance. Because of this, I digress in order to describe the National Research Council in some detail and in particular to set out how it comes to address scientific questions of national interest.

The National Academy of Sciences (NAS) was established in 1863, to serve as an entity that would "investigate, examine, experiment, and report upon any

---

99. For example the field of medicine is rife with plausible-sounding advice that has ultimately been withdrawn when the theory underlying the advice, after it is subjected to a scientific test, is found to be wrong. *See, e.g.*, Editorial, *Rethinking Hormone Therapies*, N.Y. TIMES, April 28, 2002, § 4, at 14 (questioning the effectiveness of hormone treatment for post-menopausal women).

100. The scientists of Galileo's time widely accepted the theory—published in detail by Copernicus and others long before Galileo—that the earth revolved around the sun. For a recent discussion of the acceptance of the earth-centered solar system during Galileo's lifetime, see DAVA SOBEL, GALILEO'S DAUGHTER: A HISTORICAL MEMOIR OF SCIENCE, FAITH, AND LOVE (2000).

101. Several high-profile cases exist where even worse violence to logic has been done. In these cases, juries have been allowed to decide whether certain medical treatments were pathogenic, and have decided that they were in spite of scientific testing showing that they were not. One example is that of breast implants as a cause of lupus. Even as a federal judge was working out a four billion dollar settlement in a class action, scientific results involving large numbers of women were being published showing that there was no link. James T. Rosenbaum, *Lessons from Litigation over Silicone Breast Implants: A Call for Activism by Scientists*, 276 SCI. 1524 (1997). For a sequel involving one of the few instances where a federal judge availed himself of expertise pursuant to Rule 706, see Jocelyn Kaiser, *Scientific Panel Clears Breast Implants*, 282 SCI. 1963 (1998).

subject of science or art" whenever called upon to do so by any department of the government.[102] In general, the NAS members were not government employees and the NAS itself was not a governmental agency per se.[103] The National Research Council was added in 1916, the National Academy of Engineering in 1964, and the Institute of Medicine in 1970, all ultimately under the umbrella label of National Academies. The National Research Council is the active arm of the National Academies;[104] for example, when a branch of the federal government requests the NAS to undertake a study, it is the National Research Council that establishes the group that will actually conduct the study and prepare the report. It also oversees the study and provides resources from throughout the National Academies to assist the assigned committee. Because of the immense prestige of the National Academies, these studies can draw upon any personnel resources that it feels are needed.

Participating in the just-completed National Research Council polygraph study were the Committee on National Statistics and Board on Behavioral, Cognitive, and Sensory Sciences, Units of the National Research Council's Division of Behavioral and Social Sciences and Education.[105] The working committee consisted of fourteen individuals, primarily from academia, and drawn from a diversity of technical fields, including, but not limited to, psychology, statistics, neurology, and epidemiology.[106] This attests to the universality of the scientific community. The crucial skill that an individual brings to the evaluation of an instrument, theory, or technique, is the ability to assess data directly and recognize a scientifically sound procedure, as well as to evaluate the limitations (the "error bars") of all such procedures. This is true whether the tests relate to DNA, polygraphs, parapsychology, or any other theory or procedure that adherents claim has been tested and confirmed.[107] Bolstering these skills is an ingrained distrust of all claims that certain phenomena can be observed and theories confirmed only if the right person is doing the test or if the test is conducted with only very special research equipment, or using components only available for a specific manufacturer. Of course, *someone* has to be the first to observe a new discovery, which means that the earliest reports of a discovery will always be based on specific people doing tests with specific laboratory equipment. It is precisely because of this that the modern sci-

102. An Act to Incorporate the National Academy of Sciences, Ch. 111, 12 Stat. 806 (1863).

103. *The National Academies, at* http://www.nationalacademies.org (last visited Oct. 27, 2003).

104. *See Organization of the National Research Council, at* http://www.nationalacademics.org/archives/nrcorganization.html (last visited Oct. 27, 2003).

105. POLYGRAPH AND LIE DETECTION, *supra* note 92.

106. *Id.* at 375-80.

107. The opinion in *Daubert* takes something of a detour to discuss philosopher-of-science Karl Popper and his thesis that a theory must be testable if it is scientific. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593 (1993). The word "falsifiable" is always linked with Popper's thesis because of his formulation in which this word is repeatedly used in preference to "testable." It is probably the opinion of most scientists—at least those who think at all about the Popper doctrine of falsifiability—that his revelation is completely obvious and that the only people edified by it are other philosophers and students of philosophy. Nevertheless, much ink has been devoted to this falsifiability doctrine, both in *Daubert* and in other post-*Daubert* writings on scientific evidence. As a result of Justice Blackmun's Popper references, "testability" is listed as one of the *Daubert* factors. A moment's reflection regarding the types of theories that are *not* testable shows how worthless testability is as a measure of whether particular scientific evidence is reliable. And the same follows, of course, for falsifiable.

entific method demands that in order for a discovery to be accepted at large, the first discoverers must publish their results in a peer-reviewed journal,[108] along with a methods description that is clear enough so that others who are not acquainted with the discoverers or their laboratory can understand them. The next step is for others who are following those methods to see whether they can reproduce the results. It is a sad fact that most exciting new discoveries turn out not to be sustainable (that is, believable) once this second stage of the scientific process is reached. The more exciting the discovery, the quicker the denouement, because exciting discoveries bring in many evaluators. The history of polywater, "cold fusion," N-rays, and other such earthshaking "discoveries" should heighten the wall of the courtroom against theories and techniques not accepted by the scientific community at large. In each case, the first reports were made by reputable and respected scientists (although, in the case of "cold fusion," the first report was at a press conference rather than through the accepted method of dissemination of scientific information).[109] In each case, after attempts to replicate the results by other scientists failed, these theories were rejected both by scientists working in the relevant fields and by scientists at large. Also in each case was left a small residue of true believers, scientists who could not give up an idea that seemed so exciting, and even some dishonest practices emerged to keep the idea alive. Telling is the response of one scientist to a science reporter's question at the outset of the cold fusion flurry, asking whether the whole thing was a fraud: "No, but give it two months and it will be."[110] It is this winnowing ability that usually leads to objective science discarding many of the so-called tests of validity that supporters of a technique initially tout.

In the present context it is the use of the polygraph in relation to specific instances, such as determining whether the examinee is lying in response to questions about certain well-defined events, whether he stole the money from a particular store, sold particular classified documents, or murdered his girlfriend last Friday night, that is of interest. Because of this, it may seem at first to be unfortunate that the National Research Council was charged not to evaluate polygraphs as used in such contexts, but rather to evaluate their efficacy in detecting security risks through widespread screening of applicants for sensitive positions.[111] The results of such a study would also be relevant to general pre-employment and on-the-job screening for dishonesty.

---

108. A person who does not understand "peer-reviewed" or what is meant by "peer-reviewed journal" is a person innocent of knowledge regarding how scientists work. These terms are embedded in the system by which scientists turn their research results into papers published in the professional literature, which is comprised of peer-reviewed journals. A paper submitted for publication to such a journal is farmed out by the receiving editor to two or three reviewers working in the scientific field from which the paper emanates. Based on the reports of these reviewers ("referees") the paper is accepted, accepted subject to revision, or rejected. From the language of *Daubert*, it is clear that the Court understood "peer-reviewed." Daubert v. Merrell Dow Pharm., Inc., 509 U.S. at 594. Not so for some subsequent courts, which seemed to believe that talking to one's "peers" about one's method of measurement constitute "peer review." *See, e.g.*, State v. Irving, CR-SOM-01-599 (Me. Super. Ct., Som. Cty., _____) (Marden, J.), *aff'd* 2003 ME 31, 818 A.2d 204.

109. GARY TAUBES, BAD SCIENCE: THE SHORT LIFE AND WEIRD TIMES OF COLD FUSION 104, 112 (1993).

110. *Id.* at 314.

111. POLYGRAPH AND LIE DETECTION, *supra* note 92, at 2.

The charge to the National Research Council was not to carry out its own validation tests of polygraphs (an immense undertaking, as mentioned above), but rather to evaluate the validation tests that had been done to date.[112] In other words, the Committee was to evaluate the testers and their results.

It turned out that essentially no validation tests of polygraphs for general screening purposes had ever been done.[113] Reasoning that if the polygraph was to be effective in the "soft" area of probing for general dishonesty, the Committee decided that a necessary, though probably not sufficient, condition was that it had to be effective when used in connection with specific events.[114] Consequently, the Committee set out to locate and evaluate all existing polygraph validation studies, nearly all of which related to questioning targeted persons about specific events. Out of the many thousands of articles published on polygraphs, the Committee turned up a total of 194 experiments presented as validation studies.[115]

The Committee's staff then reviewed the experiments, just retaining for study those experiments that reported examiners' conclusions regarding specific questions, and also reported the truthfulness with which the examinees had answered those questions.[116] There were further criteria applied as well, as set out in the report's Appendix G, which explains how the original 194 studies were reduced to fifty-seven for evaluation by the Committee.[117] In assessing the significance of the tests reported in the fifty-seven prior studies, the studies were *not* combined into a meta-analysis, in part because of the heterogeneity of conditions and techniques among the various tests.[118] Although meta-studies are not objectionable per se, it is very important to establish the similarity of the experiments from which data is combined. Further, some meta-studies are meaningless by their very nature. Failure to recognize the significance of the way a test or collection of tests is structured can lead to nonsensical results. Consider, for example, a reexamination of a tennis match using a meta-analysis, wherein it is argued that the loser of a set of matches was really the winner of the match because he won more individual games.

Many of the fifty-seven studies in the Committee report had more than one set of data. The Committee decided to avoid an improper weighting by selecting a single data set from each of the studies, and to lump together those data sets from

112. *Id.*

113. The Committee found only four studies aimed at measuring polygraph efficacy for the purpose of screening job applicants or current employees, and only one that was directed at screening for national security purposes. *Id.* at 3-4.

114. *Id.* at 106.

115. *Id.* at 107.

116. *Id.*

117. *Id.* at 323-39. The Committee did not require that in order to be considered, a study had to have been published in a refereed journal. *Id.* at 324. Although publication in a refereed journal is a key indicator of legitimacy for a technical paper, the Committee was not looking for indicia of reliability, but was evaluating the reported tests for reliability. *Id.* Indeed, the work of the Committee with each study was tantamount to a very high degree of refereeing.

118. *Id.* That the Committee addressed the possibility of a meta-analysis is particularly relevant to the *Daubert* discussion. In *Daubert* the proffered evidence at issue on appeal before the Ninth Circuit was a meta-study excluded at the trial court. *Daubert v. Merrell Dow Pharm., Inc.*, 951 F.2d 1128, 1129 (9th Cir. 1991). The proffered testimony was based on combining data from several independent studies of teratogenicity of the drug Benedictin in humans. *Id.* None of the individual studies showed teratogenicity, though it was asserted that when all of their individual data were combined, there was some slight suggestion that the drug was teratogenic. *Id.*

different studies that were: (1) similar in nature and (2) involved the same demographic group. This led to a total of fifty data sets for analysis.

Of the fifty data sets, thirty-four were from studies that placed examiners' assessments of individual polygraph responses into three categories: (1) indicating deception; (2) indicating no deception; and (3) inconclusive.[119] This tripartite division was important, since a great deal of the supposed "success rate" in a study depended upon how the "inconclusive" responses were interpreted. Those studies that addressed the three categories permitted the Committee to determine for itself the rates of true positives and false positives.[120] (A true positive is a correct conclusion that the examinee lied; a false positive is a conclusion of lying when none occurred. Although rates of true and false negatives could also in theory be determined under laboratory conditions, these rates have little meaning in connection with actual interrogations.) The seventeen data sets from validation tests that simply categorized examiners' assessments of polygraph responses as either indicating deception or no deception, with no "inconclusive" category, did not permit any such determination by the Committee.[121]

The Committee concluded that, based on the fifty data sets reviewed, the polygraph examiner had a somewhat greater likelihood of detecting a falsehood relating to a specific event than chance alone would indicate.[122] However, from a legal standpoint, even if that conclusion could be applied to the real universe of interrogations using polygraphs, it would be insufficient to justify the admission of polygraph evidence in a court of law, whether the standard is *Frye* or *Daubert*. Moreover, the Committee concluded that its assessment of the laboratory tests cannot be applied to the real universe of polygraph interrogations concerning specific events, because the validation studies were not representative of the field use of polygraphs.[123]

Addressing the actual question the Committee had been assigned to study, it concluded further that, given the softness of the questioning involved in generalized screening for dishonesty, there was little basis for concluding that polygraphs could detect such dishonesty with a greater-than-chance likelihood.[124] This conclusion led to the Committee's recommendation that polygraph testing for national security purposes be terminated.[125]

Though the 2003 National Research Council study is more thorough in that it reports to have examined every polygraph validation study it could locate, it does not provide conclusions differing from the earlier scientific assessments.[126] Indeed, as set out in Appendix E of the Committee's report, the National Research

119. POLYGRAPH AND LIE DETECTION, *supra* note 92, at 122.
120. *Id.*
121. *Id.*
122. *Id.* at 123.
123. *Id.* at 3.
124. *Id.* at 4.
125. *Id.*
126. *See* POLYGRAPH AND LIE DETECTION, *supra* note 92, at 212 ("Almost a century of research in scientific psychology and physiology provides little basis for the expectation that a polygraph test could have extremely high accuracy."). The NRC's conclusion is quite similar to the conclusions expressed in 1917 by the chair of the NRC's Committee on Psychology, John F. Shepard. *Id.* at 292.

Council had experience with lie detector evaluation practically from its inception.[127] Its records show that William M. Marston, the man most commonly credited with inventing the polygraph technique, contacted the Council in 1917, seeking assistance in his further work with the device.[128] Marston, who was a Harvard law student at the time, and who may also have been a graduate student, since he received a Harvard Ph.D. in psychology in 1921, suggested that his device could have wartime security applications.[129] In partial response, the Council set up a subcommittee to evaluate "methods of testing for deception" and to evaluate Marston's approach.[130] Two weeks later, Marston wired the Council, "[r]emarkable results thirty deception test under iron clad precautions letter follow" and shortly thereafter sent a letter describing the experiments and reporting his work with the Boston Municipal Court.[131] It seems that the only report from the subcommittee that is extant is in the form of the letter sent to Marston by the subcommittee's chair, John Shepard.[132] After expressing his skepticism of the blood pressure technique for monitoring truthfulness, in light of earlier work that had been reported, Shepard next suggested alternative physiological parameters to work with, but cautioned that there would still not be a one-to-one correlation between the physiological changes and the truthfulness of the subject.[133] Finally, and possibly in reaction to Marston's claims of "remarkable results" and "ironclad precautions," Shepard suggested changes that Marston might implement so as to avoid experimental bias.[134] The 2003 Committee noted the similarity of those words to what continues to be said about polygraphs more than eighty years later.[135] In the early 1920s Dr. Marston testified as an expert witness in the trial underlying *Frye*,[136] and in general had a remarkable career, one that included his creating the Wonder Woman character.[137] The Committee cites the Wonder Woman Website and her mythological lie detector described there: "The magic lasso . . . was unbreakable, infinitely stretchable, and could make all who are encircled in it tell the truth."[138]

After alluding to the decades of scientific evaluations, all of which led to the same conclusions and recommendations, the Committee questioned why there had been no policy changes or, at the very least, full-fledged validation studies launched in all of the areas of interest.[139] In light of the attitudes reflected by the DOE's reception of the report that it authorized, one might suggest that the Committee's question was a rhetorical one. The report was submitted in October 2002. In April 2003, the DOE announced that the polygraph examinations are "a tool that appears in current circumstances well-suited"[140] to security goals, and further that it "does

127. *Id.* at 291.
128. *Id.* at 292.
129. *Id.*
130. *Id.*
131. *Id.*
132. *Id.*
133. *Id.* at 292-93.
134. *Id.* at 293.
135. *Id.*
136. *Id.*
137. *Id.* at 295.
138. *Id.*
139. *See* POLYGRAPH AND LIE DETECTION, *supra* note 92, at 6.
140. Office of Counterintelligence; Polygraph Examination Regulations, 68 Fed. Reg. 17,886-02, 17,887 (April 14, 2003) (to be codified at 10 C.F.R. pt. 709).

not believe that the issues that the [National Research Council] has raised about the polygraph's accuracy are sufficient to warrant a decision by DOE to abandon it as a screening tool."[141]

In spite of the DOE's response, and in spite of the benefits that various other agencies see in the use of the polygraph, the latest NRC report lays to rest the question of whether polygraphs, as presently designed and evaluated, have general acceptance within the scientific community. They do not. The NRC report should be accepted as the final word on all of the attempts to show the polygraph's ability to detect deception, and statements that there is scientific support for their accuracy should be seen for what they are: puffing and distortion.[142]

2. The Courts

Given the uniformly hostile reception of the courts to lie detectors as the source of evidence both before and after *Daubert*, it may be said that too much space has been devoted to their discussion in a paper purporting to address changes in scientific evidence and forensic science in the wake of *Daubert*. It is submitted that this space was justified as a means of providing a case study so that courts in the future can effectively address the factor of general acceptance by the scientific community, which, for most types of scientific evidence, should be the most important of the "*Daubert* factors."

Because of the continuing rejection by scientists of polygraphs, there has been little concern that polygraph evidence would ever be admitted under the *Frye* standard. With *Daubert* perceived as loosening the "overly strict" conditions of *Frye* and, in principle, establishing a standard where anything that the trial judge concluded had been done by the scientific method could be admitted, there was concern that polygraph evidence would finally be admitted. The typical judge, well-

141. *Id.* at 17,888. Four months after spurning the NRC report it had requested, the DOE backed down a little. About to face congressional criticism over the broad screening use of the polygraph, Deputy Energy Secretary Kyle McSlarrow announced a new DOE policy that he predicted would reduce the number of persons subject to polygraph screening from 20,000 to about 4,500, "mainly in sensitive arms and intelligence posts." He added the factually incorrect statement that "[n]o one has suggested that we abandon their use," followed by the disingenuous clause "or that we hire people and entrust them with national defense information with no prior checks or reviews whatsoever." William J. Broad, *Government to Give Fewer Lie Detector Tests*, N.Y. TIMES, Sept. 5, 2003, at 17A.

142. *See, e.g.*, Brief of Amicus Curiae Committee of Concerned Social Scientists, United States v. Scheffer, 523 U.S. 303 (1997) (No. 96-1133) (definitely not to be confused with the Union of Concerned Scientists, which consists of scientists). This brief supported the Respondent, citing a 1987 military court holding that polygraphs "had reached a level of scientific reliability such that they should not be routinely excluded from court-martial proceedings." *Id.* at 2. (The subsequently promulgated Military Rule of Evidence 707 categorically prohibits polygraph results from being used in any court-martial proceeding.) *Id.* Respondent in *Scheffer* had been prevented from introducing exculpatory polygraph results at trial. United States v. Scheffer, 523 U.S. at 306. The United States Court of Appeals for the Armed Forces reversed, and the government appealed to the United States Supreme Court, arguing that the reliability and helpfulness of polygraph tests are widely questioned by the scientific community, and therefore lack general scientific acceptance. *Id.* at 307. Startlingly, the Committee of Concerned Social Scientists' brief asserts "polygraph tests are generally accepted in the scientific community as evidenced by the volume of publications in peer-reviewed scientific journals and by surveys of scientists." Amicus Curiae Brief at 3.

trained in the law but with no training in science, could very well be convinced by a trained polygraph technician that the device works and is based on the scientific method.

There have, in fact, been a few examples where the concern expressed above was realized. In *United States v. Crumby*[143] polygraph results were accepted into evidence based on the following justification:

> In general, the Court concludes that the maturation of the science of polygraphy, when properly coupled with a cautious acceptance of this science by federal courts, will lead to a fairer and more just system of criminal and civil jurisprudence. Courts must assist the trier of fact in its quest to ascertain the truth, while ensuring that the trier of fact is not unduly [misled] or prejudiced. Thus, the Court is compelled to reexamine the current state of the law based on the particular set of facts in this case, the advances in the science of polygraphy, the reasoning of *Piccinonna*, and *Daubert* and *Daubert v. Merrell Dow Pharmaceuticals* . . . ("Daubert II").[144]

The *Crumby* court also stated that the polygraph had been tested via the scientific method, has peer-reviewed publications, has a known error rate, and is "accepted by polygraphers and its use is endorsed by a number of organizations."[145]

B. Fingerprints

1. The Science

Until recently, fingerprint evidence was considered the gold standard of identification, the forensic identification technique with which all others were compared. Indeed, one way in which later identification techniques were touted (consciously or unconsciously) was by suggestive labels invoking fingerprints: as with voice spectroscopy ("voiceprints") and forensic DNA analysis ("genetic fingerprinting"). Thus, the forensic science community reacted with varying degrees of shock and disbelief to claims that fingerprint identification had never undergone the type of testing called for in *Daubert*, and that there is nowhere in the open literature where one can learn about the rates of false matches ("false positives").[146]

To those not familiar with the controversy, it seemed at first that the primary, if not sole, objection of the critics was the lack of proof of the oft-repeated statement that no two fingerprints are alike. The initial responses of those rejecting the claim that fingerprint matching did not comply with the *Daubert* standard were largely centered on this issue, which, compared to the others, is essentially a straw man. This allowed defenders of fingerprint evidence to sound reasonable in ridiculing the critics, one of the most prominent of whom was Simon Cole.[147] Indeed, some of the critics of the critics continue to lash this straw man. However, they also misconstrue the significance of the critics' claim that the error rate in finger-

143. 895 F. Supp. 1354 (D. Ariz. 1995).

144. *Id.* at 1358.

145. *Id.* at 1358, 1360.

146. *See, e.g.*, Edward Imwinkelried, *Fingerprint Science*, NAT'L L.J., Dec. 23, 2003 at B14.

147. SIMON A. COLE, SUSPECT IDENTITIES: A HISTORY OF FINGERPRINTING AND CRIMINAL IDENTIFICATION (2001).

print matching is not zero.[148] Even Henry Lee, perhaps the best known forensic scientist in the country, characterizes in a recent book the fingerprint dispute as involving no more than an argument over the standard that should be used in declaring a match.[149] Whereas there has been criticism of the varying standards or lack of standards of this type, this also is not the primary issue of the critics.

The major criticism of those claiming that at present, the technique of latent fingerprint identification does not satisfy *Daubert* is that there has never been a validation study of the technique. More specifically, they claim there has never been a scientific validation test of the rate of false positives (erroneous matches) or false negatives (failure to see a match that does exist) when partial fingerprints are compared with full rolled fingerprints of known persons.[150] Prints lifted from crime scenes normally contain only a fraction of the fingerprint of the individual who left it there. They are partial prints.[151] It does not require someone experienced in fingerprint identification to realize that, as the print fraction gets smaller, there will come a point where it does not contain enough information to support a match between the latent print and a specific known print. Stated differently, there will come a point at which the information contained in a crime-scene print is consistent with the fingerprints of more than one person.[152] Furthermore, even with fairly large partial prints, the poor quality usually inherent in latent prints recovered from a crime scene compromises the comparison. The typical response of fingerprint examiners is that they know what they are doing and know when there is not enough information in the latent print to support a match.[153] The

148. See, for example, the website of Ed German, a fingerprint examiner for the U.S. Army criminal investigation laboratory, who suggests that the critics are saying that because some examiners make errors some of the time, the whole enterprise should be discarded. Introducing an argument by analogy that has come up more than once in this very form, Mr. German states, "[m]ath is not bad science despite practitioner error." Ed German, *Regarding Recent News Articles on Fingerprint Evidence Credibility in Court, at* http://onin.com/fp/stmt_ref_articles.html (June 11, 2001).

149. HENRY C. LEE & THOMAS W. O'NEIL, CRACKING CASES: THE SCIENCE OF SOLVING CRIMES (2002).

150. The FBI reports having used its AFIS to compare every one of 50,000 rolled prints with every one of a second set of 50,000 rolled prints, where it had been documented that no overlap should exist between the sets. It reported that not a single match was found in this automated search which, by its nature does not have a match-distinguishing ability as great as that of trained human examiners. Goverment's Combined Report to the Court and Motions in Limine Concerning Fingerprint Evidence, United States v. Mitchell, 199 F. Supp. 2d 262 (E.D. Pa. 2002) (No. CR.A. 96-407-1). It is important to note the distinction between rolled prints—complete fingerprints secured under controlled conditions, usually by a fingerprint technician—and latent prints, essentially always partial prints of mediocre quality inadvertently left at a crime scene.

151. There is sworn testimony by an FBI official that the average latent print contains only about 21% of the full fingerprint. Hearing Day Two Transcript of Stephen Meager, United States v. Mitchell, 199 F. Supp. 2d 262 (E.D. Pa. 2002) (No. 96-407-1), *available at* http://www.clpex.com/Information/USvMitchell/2DaubertHearingTranscripts/US_v_Mitchell_Daubert_Hearing_Day_2.pdf (last visited Nov. 22, 2003).

152. The second basis for the criticism is the lack of uniform standards guiding fingerprint examiners in determining whether a match exists. Although fingerprint-matching has a large subjective component, there are still some objective components where quantitative standards could be imposed, such as the minimum number of matching points ("Galton points") necessary to declare a match. In contrast to most police agencies in the world, most U.S. agencies, including the FBI, do not require a minimum number.

153. *See generally* CRAIG A. COPPOCK, CONTRAST: AN INVESTIGATOR'S BASIC REFERENCE GUIDE TO FINGERPRINT IDENTIFICATION AND CONCEPTS (2001).

critics say that though that may very well be true, it is not enough, from a scientific method point of view, to simply make this claim in the absence of validation tests that are open to review by everyone.[154]

In spite of various jeremiads that the technique has been tested for over 100 years,[155] it seems undisputed to one understanding the nature of scientific testing and statistical analysis that the type of validation studies referred to above have not been done. It also seems, given the computerized Automatic Fingerprint Identification Systems (AFIS) that now exist, such tests could be commenced immediately, especially in the FBI with its huge database of prints.[156] The rate of false positives can be studied as a function of many different variables, including the examiner training, size of latent print (both in terms of area and in terms of information points it contains), quality of latent print, and the nature of the surface from which it was obtained. The beauty of having the data already available, and largely in electronic form, is that even starting with a narrowly directed study, valuable information can be obtained.

Rather than undertake such studies, many of the large police agencies and individual fingerprint examiners are taking a defensive stance, attacking the credentials of their critics and making arguments that do not address the significance of the validation tests.[157] An example of this was observed by this author at the 2002 Annual Meeting of the American Academy of Forensic Sciences in Atlanta. At a well-attended breakfast seminar, Simon Cole presented a talk on the subject of the fingerprint dispute. Seated together at a large round table close to the lectern was a group of fingerprint examiners and administrators, including a number from the FBI. After making *sotto voce* comments throughout the talk, at the conclusion of the talk, one of the people in this group asked rhetorically, "[S]uppose that in a roomful of mathematicians one of them made an error in adding up a column of numbers. Does that mean that the entire science of mathematics is wrong and should be discarded?" He concluded by lifting up two cups of coffee and challenging Cole to say how many cups were in his hands. Cole responded that the analogy was not apt. Someone from the audience then pointed out that scientists and mathematicians do not consider mathematics to be a science in the first place, and that line of discussion ended.[158]

Some responses from persons presuming to speak for fingerprint examiners do attempt to address the critics' call for validation studies. These responses in-

154. *See* COLE, *supra* note 147.

155. United States v. Llera Plaza, No. 98-362-10, 2002 WL 27305, at *11-12 (E.D. Pa. Jan. 7, 2002).

156. *See supra* note 150.

157. It is possible that police agencies and examiners, many of whom lack formal scientific training, simply do not understand what a scientific validation test is, and the fact that one does not have to be an expert in a technique in order to evaluate the tests that purport to validate that technique. This is suggested by the government's argument in *Llera Plaza*, an argument based on the conclusory assertion that fingerprinting had already been "tested" for over one hundred years by juries that had convicted defendants based on fingerprint evidence. United States v. Llera Plaza, 2002 WL 27305, at *9.

158. Although the retort that mathematics is not a science was itself a cheap shot, the fact is that even if the analogy had been made to a science, it would still have been inapt and misleading. No one is saying that fingerprint matching should be abandoned because it is not perfect. The demand is simply that the level of "perfection" be tested for.

159. *See* Andre A. Moenssens, *Is Fingerprint Identification a "Science"?*, *at* http://www.forensic-evidence.com/site/ID/ID00004_2.html (last visited Nov. 22, 2003). Moenssens is the author of several books on the subject of fingerprint evidence.

clude the observation that declaring a match between a latent print and a rolled print is not amenable to statistical analysis, or to the setting of any absolute standards. This line of response suggests further that, as far as identification goes, fingerprints are better than DNA. Unlike a DNA match, which must be accompanied by a probabilistic statement regarding the likelihood that a match exists with a second person, a fingerprint match is a definitive identification of the person identified, with 100% likelihood.[159] Apart from glossing over the minuscule probability that a forensic DNA match does not indicate that the unknown sample must have come from the known individual, this is yet another masking of the controversy, shoving it back into the false dispute about fingerprint uniqueness. The critics are not denying that a correct fingerprint match definitively establishes that the defendant left the print at the crime scene, but rather, argue that the probability relates to the likelihood that the match is correct, something that cannot be done at present because the necessary validation studies have never been done.

There are also some responsive answers to the critics' complaint that there are no standards for establishing that a match exists. In part, this relates to differences among various police agencies in the U.S. and other countries as to whether there should be a minimum quantitative standard for declaring a match. Many agencies, especially outside the U.S., do have such a standard, based on the number of Galton points (referring to ridges, bifurcations, etc.) that must be the same on the two prints being compared. Some agencies require a minimum of fourteen, others sixteen. The FBI does not have a minimum requirement, something that it dropped from its matching protocol more than fifty years ago.[160] This means that the FBI follows no minimum standard, and certainly not a quantitative one. Its primary response to this criticism is that there is a lot more to fingerprints than Galton points, and that in the end the decision as to whether a match exists is a subjective one.[161] That is, the full picture of a fingerprint includes the second order aspects of the ridges, grooves, etc., aspects such as the variation in their height and depth along their length that are impossible to quantify or to describe with precision. In this regard, fingerprint matching is like general pattern recognition, which involves elements that cannot be articulated. Consider, for example, the typical human's ability to recognize thousands of different human faces at a glance, coupled with an inability to state the reasons for the recognition apart from listing a few salient facial features common to millions of people.[162]

160. Hearing Day Two Transcript of Stephen Meager at 105, United States v. Mitchell, 199 F. Supp. 2d 262 (E.D. Pa. 2002) (No. 96-407-1), *available at* http://www.clpex.com/Information/USvMitchell/2DaubertHearingTranscripts/US_v_Mitchell_Daubert_Hearing_Day_2.pdf (last visited Nov. 22, 2003).

161. Hearing Day One Transcript of David Ashbaugh at 115, United States v. Mitchell, 199 F. Supp. 2d 262 (E.D. Pa. 2002) (No. 96-407-1), *available at* http://www.clpex.com/Information/USvMitchell/2DaubertHearingTranscripts/US_v_Mitchell_Daubert_Hearing_Day_1.pdf (last visited Nov. 22, 2003).

162. The computer programs now being developed to provide automated facial recognition have to date not come close to the human ability in this regard. They are analogous to AFIS, used by the FBI, for narrowing down the number of fingerprint records to be compared with an unknown print. The latter are based in part on class characteristics (such as tented arch, left slope loop, etc.), and present to the examiner a certain number of "presumptive" matches, which the examiner than reviews to see if there is an actual match. The number of prints turned up by AFIS depends on how many characteristics the fingerprint technician specifies for matching.

The defenders of the thesis that no validation tests need be carried out on fingerprint matching correctly point out that it was not with *Daubert* that the first questioning of fingerprints occurred. Fingerprint matching has been questioned and accepted for decades, with fingerprint technicians being cross-examined on their techniques, their credentials, and their proficiency. Presumably, if the validation studies are carried out and the expected high success rate is verified, there will be no more global attacks on the fingerprint technique. However, the case-by-case inquiries will continue, with whatever insight into the sources of potential error the validation studies provide. The alternative is to ignore the *Daubert* standard of reliability when it comes to fingerprints and other long-standing techniques and grandfather them in, creating a double standard. Attempts to allow fingerprint evidence using the *Daubert* standard have resulted in some very strange reasoning in judicial opinions, at least reasoning that seems strange to scientists and others acquainted with the scientific method.

2. *The Courts*

No defense attorney has successfully argued that expert testimony based on fingerprint evidence should be excluded because fingerprint matching in general does not satisfy the *Daubert* standard.[163]

The closest that the defense in a criminal case has come to success on a *Daubert* motion to exclude a fingerprint expert from testifying that a match existed occurred in *United States v. Llera Plaza*.[164] Based on essentially the same arguments for and against exclusion that had been raised a short time earlier in the same court,[165] Judge Pollak ruled, in response to a defense motion, that the expert fingerprint witness could only point out to the jury the similarities between the latent print from the crime scene and that of the defendant. In other words, the expert could not take the next step and declare a match; that was something for the jury to decide.[166] In granting the defense motion, Judge Pollak indicated that he knew what scientific validation studies were, and their value.[167] While taking note of the uniqueness and permanence of fingerprints, he clearly contrasted scientific validation tests of fingerprinting techniques with what the government attorneys seemed to think they were, as reflected by their statement, "'[t]he ACE-V process and the experts' conclusions have been tested empirically over a period of 100 years and in any particular case they can be tested by examination of the evidence by another expert.'"[168] Judge Pollak responded by stating:

> The second clause of this sentence seems to be arguing that, following testimony by one fingerprint examiner that a particular latent print corresponds with a particular known print, testimony by a second examiner constitutes a form of "test-

163. For a list of trials—including one in the federal court for the District of Maine—involving this issue, and a few appellate decisions, see *Legal Challenges to Fingerprints*, *at* http://onin.com/fp/daubert_links.html (last visited Sept. 30, 2003).

164. No. 98-362-10, 2002 WL 27305, at *11-12 (E.D. Pa. Jan. 7, 2002).

165. United States v. Mitchell, 199 F. Supp. 2d 262 (E.D. Pa. 2002).

166. United States v. Llera Plaza, 2002 WL 27305, at *18.

167. *Id.* at *19.

168. *Id.* at *9 (alterations in original) (quoting Government's Motion and Response at 112, United States v. Llera Plaza, No. 98-362-10 2002 WL 27305 (E.D. Pa. Jan. 7, 2002)).

> ing." However, this is not "testing" of the "theory" or the "technique" of fingerprint identification in the *Daubert* sense. With respect to "theory," the fact that a second examiner, following the same "technique" as a prior examiner, reaches the same (or, indeed, a different) result, would not seem to shed any light on the validity of the "theory" underlying that "technique." With respect to "technique"—assuming . . . that the validity of the "theory" were acknowledged—it is difficult to see that a single confirmatory examination would be adequate to validate the "technique.". . . A scientist might be disposed to require scores, or perhaps hundreds, of observations before regarding the "technique" as having been "tested."[169]

In addition, Judge Pollak noted that:

> [T]he ACE-V process and the experts' conclusions have been tested empirically over a period of 100 years—apparently refers to the fact that fingerprint identification has been a customary ingredient of trials for a century. . . . "[A]dversarial" testing in court is not, however, what the Supreme Court meant when it discussed testing as an admissibility factor.[170]

With these remarks Judge Pollak inherited a windstorm of protest. His well-publicized ruling was the first time that many realized that *Daubert* might change the whole landscape of scientific evidence. Months of legal commentary and newspaper editorials followed excoriating the decision, though not by and large from that part of the scientific community that follows what is going on in the courts. Following the granting of the defense motion, the government moved for reconsideration, and two months later Judge Pollak reversed himself.[171] In essence, he said that he had come to the realization that the fingerprint technique was not a science needing to satisfy *Daubert* in the scientific context, but rather a specialty, which had different measures of reliability.[172] Also reportedly playing a role in the change of opinion was additional material placed into evidence regarding the proficiency of FBI examiners. The ruling in favor of admitting expert testimony of fingerprint examiners thus went only to FBI examiners, for whom, in Judge Pollak's view, effective validation tests had been carried out, at least ones sufficient to establish reliability for a specialty.[173] The opinion explaining why, "[i]n short, I have changed my mind" also shows an alarming deference to the English system.[174] The fact that the witness from Scotland Yard belittled the internal proficiency test used by the FBI was far outweighed by his statement that Scotland Yard used the ACE-V approach to fingerprint matching, just as the FBI did.[175]

Although a great relief went across the land with Judge Pollak's reversal of himself, the commentary regarding his logic was not all favorable among those who took the trouble to read the opinion. In fact, to those involved in the scientific enterprise, much of the recent case law on scientific evidence and associated legal commentary seemed not just to reflect a difference of intellectual disciplines, but a completely different logic system. This system recognizes a dichotomy between "relevant evidence" and "reliable evidence," and holds that a judge can determine

169. United States v. Llera Plaza, 2002 WL 27305, at *10.

170. *Id.* (alterations in original).

171. United States v. Llera Plaza, 188 F. Supp. 2d 549, 575-76 (E.D. Pa. 2002).

172. *Id.* at 563-64.

173. *Id.* at 565-66.

174. *Id.* at 576.

175. *Id.* at 575-76. To say that one uses the ACE-V approach says nothing about one's rate of false positives.

whether testimony will be helpful to the jury regardless of whether it is reliable. Typical of the more outspoken remarks were those of David Faigman in the journal *Science*, under the heading "Science and the Law,"[176] which were directed at Judge Pollak's self-reversal and Judge Crow's decision in *United States v. Cline*:[177]

> Upon reflection, [Judge Pollak] said, he had come to the realization that fingerprint identification was not a matter of "science," it was a "specialization," and thus need not meet the rigors of the scientific method [in order] to be admitted in court.[178]
>
> . . . .
>
> In doubting the value of the scientific method as the touchstone by which expert evidence is to be evaluated, judges like Pollak and Crow fail to say what should replace it. Presumably, it is some combination of "years of personal experience" and general acceptance among members of some well-meaning guild. . . . More troubling though, it reflects a basic misunderstanding of the subject of empirical expertise. . . .
>
> Judge Crow's statement is remarkable for both its candor and its utter failure to appreciate the culture attending scientific testing of hypotheses. Science does not "exist" categorically or in some concrete encyclopedia of knowledge that passes muster by, say, some committee of the National Academies of Science. . . . Courts make a fundamental error when they try to divide the world into science and specialty categories. In truth, every expert who appears in court has "specialized" knowledge of one sort or another. At best, it is specialized knowledge based upon good applied science; at worst, it is specialized knowledge based upon "years of personal experience."[179]

In reversing the first decision, Judge Pollak was obligated to explain how these factors were now met or why they were no longer relevant. Remarkably, he stated, "I concluded in the January 7 opinion that *Daubert's* testing factor was not met, and I have found no reason to depart from that conclusion."[180] Yet, somehow, he now found that the other three factors mentioned in *Daubert*, error rate, peer review and publication, and general acceptance, were satisfied. How this was possible, without testing, is a great mystery of the decision. For him, this mystery was solved by his observation that fingerprint identification, "is not, in my judgment, itself a science."[181] He likened forensic scientists to "accountants, voca-

176. David L. Faigman, *Is Science Different for Lawyers?* Sci. July 19, 2002, at 339.

177. 188 F. Supp. 2d 1287 (D. Kan. 2002) (upholding the admission of expert testimony regarding latent fingerprint matching despite acknowledging that the efficacy of such matching has never been empirically validated).

178. Faigman, *supra* note 176, at 339.

179. *Id.* at 340.

180. United States v. Llera Plaza, 188 F. Supp. 2d 549 at 564.

181. *Id.* at 560.

tional experts, accident reconstruction experts, [and] appraisers of land or of art."[182] Forensic science was a specialty, not a science.

As is discussed in Section IV, the problem highlighted by the above passages arises from the fact, unrecognized by Judges Pollak and Crow and others, that the *Daubert* factors are not independent of one another. General acceptance follows testing with favorable outcomes and peer-reviewed publications that are evaluated by the community at large.

C. *Voiceprints*

1. The Science

Although the place of polygraphs in a review of scientific evidence is obvious, voiceprints also have a special place in the history of scientific evidence. Today, more than one state has a rule for scientific evidence enunciated in the course of a decision regarding admissibility of voiceprint testimony. Maine is one of them, through *State v. Williams*[183] in 1978, and California, through *People v. Kelly*[184] in 1976, is another. A more general factor that makes voiceprints interesting in this context is that, in contrast to polygraph evidence, the courts have always been divided on the admissibility of voiceprint evidence. (It is tempting to speculate on the contrast between *Kelly* placing California firmly in the *Frye* column as it rejected voiceprint evidence, and *Williams* expressly rejecting *Frye* in Maine as it upheld the admission of voiceprint evidence.) Also in contrast to polygraphs, some of the supporters of voiceprint evidence come from solid academic and research laboratory backgrounds.

A "voiceprint" is more accurately called a voice spectrograph (or spectrogram). It is a representation of an utterance of speech (or roar of a tiger),[185] laying out as a function of time (on the x-axis) the frequencies contained in the utterance (on the y-axis), and the loudness (indicated by the darkness of the tracing). In other words, it is a graphical representation of loudness and frequency as a function of time. The voiceprint, with the quasi-ridges and whorls that this produces, has some superficial comparison with a fingerprint. The "voiceprint" label arose either because of this or the desire of the early developers to emphasize its individualization powers by comparing it to fingerprints. Probably both.

---

182. *Id.* at 563. Interestingly, this tracks the words of Professor James Starrs, one of the defense experts, following the first opinion. He suggested that fingerprint examiners should give up the claim to science and say that they were just technicians. Adrian Cho makes note of this:

> In the meantime, James Starrs, a law professor and forensic scientist at George Washington University in Washington, D.C., says that fingerprint examiners have a better chance of satisfying the flexible *Daubert* standards by declaring fingerprint identification a form of technical expertise, similar to accident reconstruction, rather than science. But fingerprint examiners aren't willing to strike such a bargain, Wertheim says. "That's the chicken's way out," he says. "We all feel that fingerprint identification is good science."

Adrian Cho, *Fingerprinting Doesn't Hold Up as a Science in Court*, SCI., Jan. 18, 2002, at 418.

183. 388 A.2d 500 (Me. 1978).

184. 549 P.2d 1240 (Cal. 1976).

185. *See, e.g.*, HARRY HOLLIEN, FORENSIC VOICE IDENTIFICATION 118 (2002).

Unlike fingerprints, where the image reflects what is actually on a person's fingers, with the quality of the representation varying only with the care with which the fingerprints are made, voiceprints are technical constructs that take on qualitatively and quantitatively different appearances depending on how they are made. One of the key variables when one is recording sound is the bandwidth[186] of the device that converts the sound wave into an electrical signal and, in particular, the bandwidth of the filter at the input to that converter.[187] In order to get the most faithful frequency reproduction, one needs to maximize the bandwidth. Unfortunately, in an analogue to the Uncertainty Principle, the better defined the frequency, the less well-defined the time variable. The point is that the voiceprint will have a different appearance depending on the bandwidth at which it is collected.

Nevertheless, there are certain standard frequencies for making voiceprints. In the context of speech, narrow-band is on the order of 45 Hz and broad band 300 Hz.[188] A typical voiceprint would be made from a low quality recording of a telephone voice (demanding ransom, for example). At some point, that voiceprint may be compared with voiceprints made of the voices of a number of suspects uttering similar words. The use of the voiceprint for identification purposes takes place in a combined "visual/aural" setting, where the voiceprints are reviewed by a voiceprint expert while he/she is also listening to the voices of the known ones and the unknown (criminal) one.

The issue, as far as scientific evidence admission is concerned, is whether such an expert has a better record of voice identification or voice distinguishing using the voiceprints than he/she (or any layperson, for that matter) does by just listening to the voices. The consensus among scientists who are not directly involved in the voiceprint field is that they probably do not, and that they definitely do not improve their accuracy greatly by using the voiceprints as an adjunct to simply listening. There is some evidence that to the extent that voiceprints help, it is in the situation where a long time has elapsed between the making of the recording of the unknown voice and those of the suspects.[189]

Responding to the disputes about the value of voiceprints, the FBI requested the National Research Council to evaluate them in the late 1970s.[190] This was just after the California and Maine courts had dealt with them, in *Kelly* and *Williams*, respectively, decisions referenced by the Committee doing the study. As indicated by the title of the study, the Committee broadened the scope of the task from the FBI's original request. As a result of the study, the FBI chose not to introduce voiceprint evidence into court, but to continue to work with voiceprints in background investigations.

---

186. Bandwidth is traditionally understood in physics and electrical engineering to mean the range of frequencies involved. One can use it to refer, for example, to the frequency bandwidth of an audio amplifier or to the frequency bandwidth over which a radio station is allowed to transmit.

187. *See* COMM. ON EVALUATION OF SOUND SPECTROGRAMS, NAT'L RESEARCH COUNCIL, NAT'L ACADEMIES, ON THE THEORY AND PRACTICE OF VOICE IDENTIFICATION (1979) [hereinafter VOICE IDENTIFICATION]; HOLLIEN, *supra* note 185, at 188.

188. Compare this with the bandwidth of good audio amplifiers, typically on the order of 50 kHz. Although most human speech has frequencies in the low hundreds of Hz as far as the fundamental sound goes, other types of sound will, in general, include a wide range of pure frequencies.

189. HOLLIEN, *supra* note 185, at 31.

190. *See* VOICE IDENTIFICATION, *supra* note 187.

2. The Courts

The issue of whether voiceprints are acceptable scientific evidence is important because if they are, an expert witness may testify to a voice identification that would otherwise be left to a fact witness, to be evaluated ultimately by the jury. For the party wishing the identification to be accepted by the jury, one more witness—especially an expert witness—testifying in support is usually an advantage.

At present, the state courts differ widely with respect to their receptivity to expert testimony based on voiceprint. Interestingly, most of these cases came down shortly before the National Research Council report. Of the states that have addressed the issue, eight upheld voiceprint admission, under a variety of standards,[191] and nine rejected voiceprints as not having adequate indicia of reliability.[192] Not surprisingly, Wisconsin was in agreement with Maine in admitting voiceprints. All but one of these seventeen decisions were pre-*Daubert*, the exception being Alaska, which, in 1999, found voiceprints acceptable, in the case that expressly switched Alaska from a strong *Frye* state to a *Daubert* state. [193]

Of the eight states admitting voiceprints, four did so after the National Research Council study on voice identification was published. Of the nine states rejecting voiceprints, five did so after the study was published. So, no pattern there either. This type of forensic evidence can be expected to produce volumes of briefs in upcoming years. Again, it is noted that only one state has addressed the question since *Daubert*.

Although most of the discussion of voiceprints and the courts has been focused on state courts, federal courts have also addressed this issue. For example, the Fifth Circuit weighed in on this topic in *United States v.* Drones:[194]

> At the time of Drones's trial, four of our sister circuits had upheld the admissibility of spectrographic evidence.... *However, all of these cases were decided before* Daubert v. Merrell Dow Pharmaceuticals. . . . No federal case decided post-*Daubert* has considered the admissibility of expert voice identification testimony.
>
> In addition to the fact that the state of the law concerning expert voice identification was ambiguous, the expert testimony presented at the evidentiary hearing demonstrates that spectrographic analysis is—and was at the time of Drones's trial—of questionable scientific validity. Most notably, at the hearing, Koenig testified that there is no proven theoretical basis for the basic underlying premise that one person's voice is truly unique and therefore identifiable. He further stated that this has resulted in a precipitous drop in the number of expert practi-

191. State v. Coon, 974 P.2d 386 (Alaska 1999); Alea v. State, 265 So. 2d 96 (Fla. 1972); State v. Williams 388 A.2d 500 (Me. 1978); Commonwealth v. Lyhus, 327 N.E.2d 671 (Mass. 1975); State *ex rel.* Trimble v. Hedman, 192 N.W.2d 432 (Minn. 1971); State v. Williams, 446 N.E.2d 444 (Ohio 1983); State v. Wheeler, 496 A.2d 1382 (R.I. 1985); State v. Kendley, 433 N.W.2d 674 (Wis. 1988).

192. State v. Gortarez, 686 P.2d 1224 (Ariz. 1984); People v. Kelly, 549 P.2d 1240 (Cal. 1976); Cornet v. State, 450 N.E. 2d 498 (Ind. 1983); State v. Free, 493 So. 2d 781 (La. 1986); Reed v. State, 391 A.2d 364 (Md. 1977); People v. Trobey, 257 N.W.2d 537 (Mich. 1977); Windmere Inc. v. Int'l Ins. Co., 522 A.2d 405 (N.J. 1987); People v. Jeter, 600 N.E.2d 214 (N.Y. 1992); Commonwealth v. Topa, 369 A.2d 1277 (Pa. 1977).

193. State v. Coon, 974 P.2d 386 (Alaska 1999).

194. 218 F.3d 496 (5th Cir. 2000).

> tioners over the past few decades, from fifty to sixty practitioners in the 1970's to roughly a dozen experts at the time of Drones's trial. While Cain testified that expert voice identification testimony has been used extensively in state and federal courts over the past thirty years, he also testified that he did not know if spectrographic evidence was widely accepted by the relevant scientific community. He also acknowledged that numerous factors—including a defendant's ability to disguise his own voice—could affect the reliability of expert analysis.[195]

The message is that the federal courts have yet to be heard from on voiceprints in the post-*Daubert* era. On the one hand one could infer that if voiceprints were admitted under the *Frye* standard, they would also be admitted under the *Daubert* standard. Countering this, however, is the fact that voiceprints have gradually fallen out of favor among the scientific community. This may, in fact, be an understatement. Consider:

> [T]he use of "voiceprints" exploded upon the American scene in the 1960s. It took years before the weight of the relevant research (plus court testimony by scientists) demonstrated the harm "voiceprints" were bringing to law enforcement and the courts. The misuses of the procedure now are obvious, as is the damage that has been done by their use.... "[V]oiceprints"... have become but a "footprint" in history.[196]

D. Handwriting Identification

1. The Science

With respect to identification techniques, the *Daubert* bell tolled for handwriting matches before it did so for fingerprinting and, so far, with more far-reaching results. Probably this was due to the fact that while we all have an ability to recognize handwriting—again a pattern-recognition skill—and in fact use this ability throughout our lifetimes, it is the rare person who could recognize his own fingerprint. Thus, the suggestion that experts in handwriting comparison might not have anything to offer that a layperson did not already have seemed reasonable.

It is true that persons who study handwriting have a formalism that is available for characterizing a specimen of handwriting when called upon to compare with another specimen, and may be more alert to superficial variations that occur in one's writing with time. The question being asked is whether the expert then can go beyond pointing out similarities and show how certain differences may not be sufficient to preclude the conclusion that two specimens were written by the same person. This is exactly the same question that was posed in Judge Pollak's two opinions regarding fingerprint experts.[197]

---

195. *Id.* at 503-04 (citations omitted) (emphasis added).

196. HOLLIEN, *supra* note 185, at 25.

197. *See supra* Part III.B.

In some ways, however, the attack on the expertise of handwriting identification has anticipated the pattern of attack on latent fingerprint identification. Most of the motions to exclude the handwriting expert from testifying as to the identity of the writer have been rejected, and the rejections upheld on appeal. On the other hand, there have been some courts that have accepted the arguments (1) that there has never been any study to show the uniqueness over time of handwriting patterns; and (2) that an expert in the field is no better at identifying unknown handwriting specimens than a layperson. As a consequence, there are efforts on many fronts to provide the data necessary to address those questions or to remedy the deficiencies. The FBI is conducting a study aimed at a statistical analysis of handwriting characteristics, and has sought volunteers with flyers which said:

> The validity of handwriting comparisons has been challenged in court under *Daubert* since 1995. Due to the scarcity of published research in this area, the forensic document community has struggled to meet these requirements. This collection of handwriting samples [being solicited] will provide a sound basis for judicial decisions regarding the admissibility of this type of evidence.[198]

People who answered the call were asked to produce *seriatim* five cursive copies of a paragraph followed by five printed copies of the same paragraph, using a total of ten pieces of paper. Volunteers had been told that this would take as long as an hour, but most people finished in about half that time.[199]

2. *The Courts*

Typical of the small number of decisions rejecting testimony by handwriting experts is *United States v. Hidalgo*.[200] The court held that in light of *Daubert*

198. This was written on a flyer distributed at the 2002 Annual Meeting of the AAFS. Note also in this regard the recent paper reporting on the use of computer algorithms to analyze the characteristics of handwriting samples from 1500 volunteers. Two types of comparisons were tested. The first involved the use of a computer algorithm to match an "unknown" writing with one of a set of known writings containing a contribution from the "unknown" writer. Correct identification was made in approximately 98% of the cases. The second task was to determine whether two specific documents were written by the same person (with no effort to identify that person). This task resulted in about 96% accuracy. Impressive as these results seem with respect to the potential for automated handwriting identification, its significance is undercut by the size and number of the handwriting samples used. Each volunteer copied a 156-word passage three times. Not surprisingly, the accuracy in identification increased with the size of the specimens compared, and decreased with the number of known samples being compared. Thus, the 98% accuracy figure was the result of performing the identification task in which a sample of only two writers was used, each of whom had provided three writing samples. One of these six samples was randomly selected as the "questioned" writing specimen, and compared against the remaining five. When the task conditions were changed so that a smaller portion of the questioned specimens was compared against a larger pool of possible writers, the accuracy decreased considerably. The accuracy of the second task (that of determining whether two documents were written by the same person) were similarly dependent upon the size of the writing specimen that was analyzed. While this study does not provide validation of the performance of human document examiners, it does take a much-needed step toward establishing a quantitative method of classifying handwriting characteristics. S. N. Srihari et al., *Individuality of Handwriting*, 47 J. FORENSIC SCI. 856 (2002).

199. April A. Yergin, personal communication. Ms. Yergin was one of the volunteers, and commented that the effort tended to leave volunteers with sore hands, suggesting that the appearance of the writing may have varied just within the hour or so of the tests.

200. 229 F. Supp. 2d 961 (D. Ariz. 2002).

there was insufficient scientific backing for the proposition that each person's handwriting was unique.[201] For this reason, the court excluded the proffered expert from identifying key handwriting as that of the defendant.[202] The court apparently did not address the more interesting question, namely whether the expert was better than a layperson in comparing two handwriting specimens. This part of the question did not bother the court, which was only concerned with the uniqueness issue.

E. Bullet "Fingerprinting"

1. The Science

Bullet "fingerprinting" by trace-element analysis is very interesting in the present context. It seems to be a technique that has been used repeatedly (subjected to the "adversarial testing" of a trial) without ever having been subjected to a validation test. Now, it appears that, at the very least, it has been overemphasized as a means of identification.

For more than thirty years, it had been thought and taught by police investigators and particularly by the FBI that every batch of lead bullets that is manufactured has a slightly different chemical composition from every other batch, and that that difference is measurable. This meant that by carrying out trace-element analysis on a bullet found at a crime scene, a bullet "fingerprint" could be developed for comparison with bullets in the possession of a suspect.[203] In recent years, the profile has been based on the concentrations (usually in parts per million) of six elements: antimony, tin, arsenic, copper, bismuth, and silver.[204]

The goal of prosecutors in using the forensic evidence developed from trace-element analysis of bullets is to be able to emphasize the unlikelihood that the match in profiles between the crime scene bullet and the bullets in the defendant's possession is purely coincidental. The legitimacy of this emphasis is, in part, inversely proportional to the number of bullets coming from a batch of lead that does not display a significant change in the trace-element profile. This theory can be rendered completely without value if the same profile exists in more than one batch.

The history of the trace-element analysis technique over the past three decades has included expert testimony from FBI agents that a given profile is limited to bullets in a single box of cartridges[205] or were all manufactured on or about the same day.[206] In a pattern that is frustratingly common, this type of testimony

---

201. *Id.* at 967-68.

202. *Id.*

203. Charles Piller & Robin Mejia, *Science Casts Doubt on FBI's Bullet Evidence*, L.A. TIMES, Feb. 3, 2003, at A1.

204. *See, e.g.*, Robert D. Koons & Diana M. Grant, *Compositional Variation in Bullet Lead Manufacture*, 47 J. FORENSIC SCI. 950 (2002); Erik Randich et al., *A Metallurgical Review of the Interpretation of Bullet Lead Compositional Analysis*, 127 FORENSIC SCI. INT'L 174 (2002); Ernest R. Peele et al., *Comparison of Bullets Using the Elemental Composition of the Lead Component*, PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON THE FORENSIC ASPECTS OF TRACE EVIDENCE, June 24-28, 1991, at 57.

205. Randich, *supra* note 204, at 175.

206. State v. Behn, 753 A.2d 1153 (N.J. 2000).

continued to be elicited and remained instrumental in convictions, even after articles began to appear in refereed forensic science literature indicating that it represented a gross exaggeration.[207] Furthermore, long before such articles began appearing, certain FBI crime laboratory scientists had their doubts regarding the individualizing power of the trace-element analysis technique.[208]

Although it could be argued technically that first validation tests were not of sufficient statistical weight to cause the technique to be abandoned abruptly, they should have led to more caution in testifying to the degree of individualization implied by the tests. Then, in 2002, a blockbuster of a paper on the subject was published that had, as one of its four authors, a scientist recently retired from the FBI.[209] This paper, *A Metallurgical Review of the Interpretation of Bullet Lead Compositional Analysis* (referred to here as RDMT, after the initials of its authors' surnames), first recounts the history of the trace-element method and the fact that there had never been any validation studies. The authors note:

> [One expert] has concluded from his study of a small number of bullets that in general bullet lead analysis does not generate individualizing information. However other forensic examiners have continued to conclude that positive association can be made between bullets associated with a suspect and those from a crime scene. *The only published rationale* purported to provide a foundation for these conclusions is that "if two bullets are produced from the same homogeneous source of lead, then they will have analytically indistinguishable composition."[210]

The RDMT then describes the authors' study of the chemical composition of many years of lead smelting to produce material for ammunition.[211] They conclude from these studies that bullets produced years apart could have trace-element profiles indistinguishable from one another and that bullets from the same "source" of lead could have profiles distinguishable from one another.[212] In addition, they discuss the entire process of ammunition manufacture, from the preparation of the lead alloy to the final production of bullets, and suggest that this shows that the entire premise of the individualization of bullets and the idea of bullet "fingerprints" was misbegotten in the first place. They conclude:

> Based on the inhomogeneities observed in lead "sources," and of numerous demonstrated instances of multiple sources that are analytically indistinguishable, it is our conclusion that the most positive opinion that can be rendered from data that show that two or more bullets or fragments are analytically indistinguishable is that they "could have," or that "it is possible" that they had, a common source. Our literature review and research suggest that no valid statistical probability or likelihood can be attached to a conclusion of "same source of molten lead" or "same box of ammunition," even if pressed by the proponent of the evidence or the court.[213]

One might expect that the RDMT validation study will be criticized by advocates of the technique on the grounds that its authors did not make any measure-

---

207. *See, e.g.*, Raymond O. Keto, *Analysis and Comparison of Bullet Leads by Inductively-Coupled Plasma Mass Spectrometry*, 44 J. FORENSIC SCI. 1020 (1999).

208. Piller & Mejia, *supra* note 203, at A1.

209. Randich et al., *supra* note 204.

210. *Id.* at 175 (emphasis added) (quoting Peele et al., *supra* note 204, at 57-68).

211. *Id.* at 176-79.

212. *Id.* at 176.

213. *Id.* at 191.

ments themselves,[214] and that, in any event, data on the source lead may not be dispositive of the issue, since the bullet manufacturing process itself may introduce inhomogeneities with individualizing potential. Indeed, three months following the publication of the RDMT paper, a report coming from the FBI's Forensic Science Research Unit appeared in the Journal of Forensic Science asserting that as a particular batch ("source") of lead proceeds through the multi-step bullet-manufacturing process.[215] It stated:

> the size of an individual homogeneous melt of lead decreases as more distinct compositions are formed as a result of remelting and mixing of sources, including lead scrap . . . [leading to] at least 10 compositionally distinguishable groups of bullet wire in a 19.7-h period. The largest group could potentially be used to produce a maximum of 1.3 million compositionally indistinguishable 40 grain bullets.[216]

While different profiles can arise from the same batch of starting lead, these findings do suggest that there is a potential for some individualization because of individual profiles that are constant across small portions of the batch at the conclusion of manufacturing. Thus, to the extent that these findings can be generalized, it is suggested that a statement, modified from what has been testified to in the past, may be made about individualizing of bullets by their trace-element profile. However, this study does not address the possibility of bullets with indistinguishable profiles coming from completely different original batches.

The FBI has recently requested that the National Research Council undertake a detailed evaluation of the trace- element analysis technique.[217] The study, titled "Scientific Assessment of Bullet Lead Elemental Composition Comparison," is described as follows:

> An ad hoc committee administered by the Board on Chemical Sciences and Technology will assess the validity of the scientific basis for the use of elemental composition determination to compare lead alloy-based items of evidence. The following three areas will be addressed:
>
> >*Analytical method.* Is the method analytically sound? What are the relative merits of the methods currently available? Is the selection of elements used as comparison parameters appropriate? Can additional useful information be gained by measurement of isotopic compositions?
>
> >*Statistics for comparison.* Are the statistical tests used to compare two samples appropriate? Can known variations in compositions introduced in manufacturing processes be used to model specimen groupings and provide improved comparison criteria?
>
> >*Interpretation issues.* What are the appropriate statements that can be made to assist the requester [that is, the FBI] in interpreting the results of compositional bullet lead comparison, both for indistinguishable and distinguishable compositions? Can significance statements be modi-

214. In part, this was because of the ready availability of data regarding trace-element concentrations in the sources. They note that not all of the trace metals found in the bullet profile are impurities, but rather that some are deliberately added in controlled amounts. *Id.* at 177.

215. Koons & Grant, *supra* note 204.

216. *Id.* at 950.

217. *The Current Projects System, The Nat'l Academies, at* http:www.nas.edu/webcr.nsf/ProjectScopeDisplay/BCST-L-02-03-A? OpenDocument (last visited Oct. 14, 2003).

> fied to include effects of such factors as the analytical technique, manufacturing process, comparison criteria, specimen history, and legal requirements?[218]

This sounds like what is needed to determine whether there should be "general scientific acceptance" of this technique. Depending on the outcome of the study, there may be many petitions for post-conviction relief, so effective has the trace-element analysis seemed to be in producing guilty verdicts.

2. *The Courts*

It is obviously too soon for the latest findings to be reflected in the courts, let alone in appellate decisions. However, consider *State v. Noel*,[219] a 1997 New Jersey appellate decision as an indication of how the trace-element analysis technique may have been used beyond its known range of validity. In this particular case, it was not the expert witness who was criticized for misstating or exaggerating the individualizing ability of the trace-element analysis, but rather the prosecutor, for the manner in which he characterized his expert's testimony.[220] The testimony *in toto* indicated that hundreds of thousands of 9mm bullets could have the same profile, having come from the same batch of lead alloy used in the bullet manufacture.[221] The court found that without being able to testify to the total number of bullets that would have the same profile, and the patterns of distribution of the bullets from the particular batch, there was inadequate foundation for the expert's testimony to be more probative than prejudicial.[222] The court criticized the prosecution's closing statement, where he seemed to imply (with phrases like "unique as a snow flake," harking back to the analogy made of actual fingerprints) that the similarity of the profile that was found had more significance than it did.[223]

The most interesting facts in *Noel* come from the dissenting judge's opinion, which revealed that there were nine bullets in the defendant's possession and six from the crime scene that were tested, and that these fifteen bullets were found to display five different profiles.[224] One would think that this distribution in itself would have raised questions about the basic premise of bullet matching using trace-element analysis.

Although the verdict was reversed in *Noel* on grounds related to the trace-element analysis, the *Noel* court was satisfied that there was general scientific acceptance of the basic technique.[225] This was five years before it became clear that the technique had never been validated in a manner having statistical significance. The court cited cases from other jurisdictions over the previous twenty-four years that had accepted the trace-element analysis as acceptable scientific evidence.[226]

---

218. *The Current Projects System*, *at* http://www.nas.edu/weber.nsf/ProjectScopeDisplay/BCST-L-02-03-A?OpenDocument.

219. 697 A.2d 157 (N.J. Super. Ct. App. Div. 1997).

220. *Id.* at 161.

221. *Id.*

222. *Id.* at 164-65.

223. *Id.* at 165.

224. *Id.* at 167.

225. *Id.* at 162.

226. *Id.*

In June 2003, an FBI chemist pled guilty to perjury in pretrial testimony offered in connection with a Kentucky murder trial in 2002.[227] In particular, she stated that the bullets found in the possession of the defendant and at the murder scene were manufactured in a process that resulted in no more than 280,000 bullets per batch with the same profile and the same bullet fingerprint.[228] When a witness from the bullet manufacturer contradicted that statement, the chemist admitted that she had lied, and that the number in question should have been in the tens of millions.[229] She was quoted as blaming her conduct on a sense of crisis in her work, fed by "new and repeated challenges to the validity of the science associated with bullet lead comparison analysis."[230] It is reasonable to conclude that the "new and repeated challenges" arose because of *Daubert*.

IV. CONCLUSIONS AND RECOMMENDATIONS: *DAUBERT* FACTORS REDUCED TO THEIR WORKABLE ESSENCE ARE ESSENTIALLY THE *FRYE* STANDARD

The Court, in *Daubert v. Merrell Dow Pharm., Inc.*,[231] was careful to say that the measure of admissibility of expert testimony was reliability and not the degree to which it matched up with any particular list of reliability indicia. In *Kumho Tire Co. v. Carmichael*,[232] the Court explicitly stated that Rule 702 applied to all expert opinion testimony, regardless of whether it was scientific. This sent the message that, as the nature of the testimony departed further from what might be considered a scientific or technical field, the types of things considered in assessing reliability could be very different from the factors set out in *Daubert*. However, this Article is about expert testimony relating to scientific evidence. When it comes to evaluating scientific evidence under *Daubert*, the strong tendency of the courts is to "check off" the *Daubert* factors in the same manner that courts dealing with likelihood of confusion in trademark infringement cases tick off the *DuPont* factors.[233] It is therefore important for a court to look more closely at the *Daubert* factors, which would reveal that expert scientific testimony does not always meet this standard.

When proffered scientific evidence is implicitly or explicitly awarded "merit points" based on how it satisfies each of the *Daubert* factors, the balance is shifted toward admitting unreliable testimony into evidence, the degree of shift depending on exactly how the factors are articulated. To see why this is so, refer to the Section III listing of *Daubert* factors, a configuration typical of those appearing in "*Daubert* hearing" decisions. With a little reflection, one can see that the first of those factors lacks any practical significance and therefore provides no guidance at all for evaluating proffered scientific evidence. As will be set out below, two other factors are effectively redundant, and therefore lead to double counting if considered separately, compounding the error of considering the first factor at all. In sum, three of the *Daubert* factors are logically tantamount to a single one, namely (c) whether the technique has a reasonable error rate. After the following discussion, a suggested form for the combined factor will be given.

227. Mark Pitsch, *Ex-FBI Scientist Pleads Guilty*, COURIER-J. (Louisville), June 18, 2003 § B.
228. Piller & Mejia, *supra* note 203.
229. *Id.*
230. *Id.*
231. 509 U.S. 579 (1993).
232. 526 U.S. 137 (1999).
233. *In re* E. I. DuPont DeNemours & Co., 476 F.2d 1357, 1361 (C.C.P.A. 1973).

Factor (a), whether the technique is testable and tested, is nearly always cited by courts explaining their *Daubert* analyses, is without practical significance. To award a technique points just because it is testable is absurd. The fact that testability is even alluded to in *Daubert* probably results from the deep philosophical waters through which Justice Blackmun swam beside Karl Popper. Indeed, much of the discussion in *Daubert* going to just what is "science" reflects such a swim.[234] The message of that discussion is that before one can evaluate the reliability of something proffered as scientific evidence, one must check to see whether it meets the definition of science as popularized by the now largely discredited theories of Popper regarding how science is "done."[235] Those contributions of Popper to the general discourse that have *not* been discredited are obvious to most persons who reflect on them and were recently summed up by Martin Gardner, who said, "I believe that Popper's reputation was based mainly on his persistent but misguided efforts to restate common-sense views in a novel language that is rapidly becoming out of fashion."[236] Nevertheless, Popper *did* contribute to many non-scientists the idea that the essence of a scientific theory was that it was testable, even though he expressed this by saying that it had to be "falsifiable."

One quickly finds, when devising examples that are *not* testable, that applying the "testability" criterion to proffered scientific testimony is not a very good use of one's time. One example of a non-testable theory would be one that depended on variables that could not be observed directly or indirectly (sometimes called "hidden variables"). For example, in place of the present atomic theory based on quantum mechanics and various interaction forces between sub-atomic particles, one might propose that all observed atomic behavior was due to tiny, tiny creatures, forever undetectable by any means, controlling the electrons and nuclei. For every objection that could be raised (such as the impossibility of sentient creatures smaller than the hydrogen nucleus), an ad hoc, untestable explanation would be offered.

It is, in fact, quite difficult to think of a theory of *anything* that is not testable in a statistically acceptable manner. In contrast, it is easy to list really bizarre theories that are testable, and therefore scientific. In 1955, the list included the theory that the moon was covered with miles-thick dust which would swallow anything landing on its surface.[237] It could have also included the theory that the moon was shaped like a giant carrot, with its pointy end always facing away from

234. In fairness to Justice Blackmun, it is noted that more than one of the *amici curiae* indulged themselves in philosophical discussions that went far astray. Therein undoubtedly lay Justice Blackmun's infatuation with Popper and his theories.

235. Popper became obsessed with the idea that scientific work is an attempt to "falsify" scientific theories. Essentially no scientific work is so directed, though some results may be inconsistent with a theory and if enough inconsistencies arise and another theory is available, the first theory will be discarded. *See, e.g.*, KUHN, *supra* note 35.

236. Martin Gardner, *A Skeptical Look At Karl Popper*, SKEPTICAL INQUIRER, No. 4, 2001, at 13. The major complaint about Popper, which was raised decades ago, is that he claimed that science consisted of attempts to falsify conjectures, whereas normal scientific work assumed that results consistent with theory will be obtained. The falsification that does occur arises from the normal course of work rather than an attempt to falsify. Furthermore, as Kuhn has pointed out, a theory is not simply "falsified" even if a valid experimental result is inconsistent with that theory. *See generally* KUHN, *supra* note 35.

237. This was a theory advanced by the respected astrophysicist Gerard Kuiper (1905-1973), who in addition to publishing it in the professional literature, presented it in a symposium at the University of Chicago in 1957, attended by this author.

the earth. Continuing with the moon, another testable theory yet to be tested is that the moon has green cheese at its core or that it contains an underground civilization of proto-humans. The point is that requiring a theory to be "testable" does not really eliminate anything; in particular, it is highly unlikely that it would eliminate anything proposed as scientific testimony. Testability is a necessary, but not sufficient, condition, and to list it as one of a small number of guides to reliability is to dilute the inquiry into reliability. Furthermore, although the factor has been listed above as "testable *and* has been tested," many courts devote text to just the simple "testable" factor.

Of course, the phrase "and has been tested" does not add anything to the value of the factor as a measure of reliability. So what if it has been tested! It might have failed, given ambiguous results, or shown itself to be in other ways undependable. And yet, again, one sees judicial decisions awarding admissibility points to a forensic technique because it has been *tested*. It is submitted that the only testing-related attribute that contributes anything to gauging the reliability of a theory or technique is that of having been adequately tested and found to be accurate. Depending on the particular type of testimony involved, this can be stated in any one of a number of ways.

Factor (d), the "standards" factor is also, in general, superfluous. In order for a technique to be tested in a way that is significant, there must be some standard manner of performing the technique. It would, however, be useful, in connection with the technique's particular application that is before the court, to ask whether the applicable standards had been complied with.

In terms of scientific and legal logic, three of the five "factors" listed above collapse into a single factor stated in terms of a value assessment. In keeping with the form used by *Daubert*, it can be stated as:

Have statistically valid tests of the theory
(technique, etc.) shown it to be accurate?

Note that the plural "tests" is used. As far as scientists are concerned, if the answer to that question is in the negative, that is as far as one goes.

How do scientists learn about new theories and techniques and their validations? By the refereed scientific literature. This does not mean that new theories and techniques published in the refereed literature (Factor (b)) automatically become accepted.[238] It is only when the scientific community can conclude that the answer to the question regarding accuracy is in the affirmative that this acceptance occurs. Therefore, though having one's theory published in the refereed literature indicates that it is more than just "junk science" dreamed up for the trial in question, publication by itself does not establish reliability.

In summary, the *Daubert* factors, when applied logically to proffered forensic testimony, reduce to the following:

Does the technique on which the testimony is based
have the general acceptance of the scientific community?

In cases where the technique is of such a nature that the breadth of the scientific community familiar with it may be small, one may add:

---

238. It generally does mean, however, that the discoveries are worth reading about. The goal of the referees is to eliminate the articles that, as Pauli famously said, are so bad that they are not even wrong.

Have statistically valid tests of the theory (technique, etc.) shown it to be accurate?

Expert testimony from a witness educated in statistics (there are more of those around than one might think) would generally be all that is necessary to resolve the latter question.

There has been a great deal of discussion in the legal and scientific communities since *Daubert* about the use of "court's experts" under Federal Rule of Evidence 706. Although there has been a marked reluctance by both the plaintiffs' bar and the defense bar to see this suggestion get very far, there may be much less reluctance to allow the testimony of a Rule 706 expert that is limited to evaluating the statistical significance of proffered validation tests.[239]

Scientific evaluations of the polygraph show why one of the major complaints about the "general acceptance" guideline of *Daubert* is invalid. The complaint is that one cannot assess general acceptance without first determining what the appropriate scientific community is, and that whether or not the technique in question is found to have general acceptance depends on which community one looks to. This is missing the point; it is usually the entire scientific community to which one can look for the existence of consensus regarding a particular theory or technique. The National Research Council studies identified in this Article provide one illustration of why this is so.[240] Notwithstanding the fact that science is a highly complex enterprise comprised of specialized working communities, whose members are generally unable to converse with one another about the substance of their respective work, there remains a unifying bond among scientists, a universal language. That bond and language are strengthened by their scientific approach to the physical world, an approach that enables most scientists to understand and assess validation studies regardless of the specialty to which the studies refer. Such assessment is, in fact, probably what scientists as a group are best at.

Furthermore, not only are they able to detect the weak element or fatal failing in validation studies underlying bootless claims and theories, most can articulate in non-scientific terms the rationale for their conclusions. The reason why this can be done across community borders is because the evaluation is basically procedural, just as it is, for example, when a jurist evaluates the fairness of a trial or other legal procedure. Consequently, "general acceptance" really means general acceptance by the entire scientific community, an acceptance that can be gauged by existing literature or by requesting a group such as the National Research Council to determine it.

Finally, in view of the stakes that are presented by the exclusion or admission of forensic testimony, and the interpretation that is given to that which *is* admitted, it is difficult, if not impossible, to see how a criminal defendant can receive a fair trial, or a criminal defense attorney prepare a reasonable defense, without there being forensic resources available to the defense. So important is this point, as can be seen from just a review of the history of bullet "fingerprint" analysis, that it would seem that no criminal trial should be permitted to go forward without the defense having acquired broad technical assistance in dealing with the prosecution's

239. A program to assist in Rule 706 experts has been initiated by the AAAS, assisted by the AAFS and other scientific organizations under the sponsorship of the NIJ. However, three years into its existence, the provided experts number in the single digits.

240. *See supra* Part III.A.

expert witnesses and forensic sciences. There is little doubt that trial courts in Maine already strongly discourage defense attorneys from proceeding without such assistance.[241] Defense attorneys should be barred from doing so, and sanctioned for failure to be ready to proceed if they have not obtained such assistance by the day scheduled for trial.

241. See, for example, the transcript of the pre-trial hearing in *State v. Crawford*, CR-00397 (Me. Super. Ct., Pen. Cty.), in which the defendant was charged with arson. The trial judge repeatedly queried the defense's decision not to consult a fire expert before the case went to trial. The prosecutor also questioned this decision, citing the likelihood that a guilty verdict would be appealed on this ground, given that the prosecution case turned largely on testimony by the prosecution's own fire investigators. Subsequently, after serving a term in prison, the defendant petitioned for post-conviction relief based on his loss of opportunity to have a forensic expert. A decision, issued on September 2, 2003, denied the petition based on the court's finding that Mr. Crawford had knowingly waived his right to expert assistance in the pre-trial exchange in which he personally took part. Crawford v. State, CR-02-882 (Me. Super. Ct., Pen. Cty., Sept. 2, 2003) (Hjelm, J.).