A New Era in the Evolution of Scientific Evidence - A Primer on Evaluating the Weight of Scientific Evidence

Edward J. Imwinkelried
A NEW ERA IN THE EVOLUTION OF SCIENTIFIC EVIDENCE—A PRIMER ON EVALUATING THE WEIGHT OF SCIENTIFIC EVIDENCE*

EDWARD J. IMWINKELRIED**

Modernly, scientific proof is one of the chief types of evidence in criminal cases. In a recent survey of judges and attorneys by the National Center for State Courts, forty-four percent of those responding stated that they encountered scientific evidence in at least thirty percent of their cases. There are numerous scientific evidence texts tailored for criminal practitioners, and virtually every seminar on criminal practice includes a presentation on scientific proof. In the words of one experienced prosecutor, scientific proof has become “the backbone of every circumstantial evidence case.”

To appreciate how scientific evidence gained this critical role, we must review the stages in the evolution of scientific evidence. Generalization about legal development is always dangerous, but we can discern three different eras in the modern history of scientific evidence.

At the beginning of the 1960's, scientific evidence was relatively unimportant. Prosecutors relied primarily on physical evidence and lay testimony, especially the testimony of eyewitnesses. For their part, defense counsel responded in kind. However, in that same decade the Warren Court began fashioning the fourth, fifth, sixth, and...

---

* This article is based in part on the editorial commentary the author prepared for SCIENTIFIC AND EXPERT EVIDENCE (E. Imwinkelried ed. 1981), recently published by the Practicing Law Institute of New York.

** Professor of Law, Washington University; B.A., 1967; J.D., 1969, University of San Francisco.


and sixth amendment exclusionary rules. Those rules restricted the admissibility of the physical evidence and lay testimony that the prosecution had relied so heavily upon.

In the minds of many police and prosecutors, the Warren Court decisions created an "evidence void." Police agencies developed a new approach to investigation and placed much greater emphasis on forensic techniques. Similarly, prosecutors felt compelled to resort to scientific evidence much more frequently than they had in the past. Although many defense counsel would deny that the Warren Court precedents drastically limited prosecution use of confessions or physical evidence, it is clear that prosecutors held that perception. Whether or not that perception was correct, the perception motivated prosecutors to employ scientific evidence much more often.

Predictably, the number of cases involving scientific evidence increased dramatically during the 1970's. During the 1970's, the courts were overwhelmed with such strange forensic techniques as atomic absorption (AA), forward looking infrared (FLIR), neutron activation analysis (NAA), nuclear magnetic resonance (NMR), and scanning electron microscopy (SEM). As the prosecutors offered these novel types of scientific evidence, they forced the courts to learn a whole new vocabulary.

The defense bar countered by resurrecting the Frye test to block the admission of the prosecution's scientific proof. Frye had been

---

8. Giannelli, supra note 1, at 1199-1200.
decided a half century earlier in the 1920's. In Frye, the appellate court sustained the trial judge's ruling excluding a forerunner of the polygraph, systolic blood pressure evidence. Frye is a landmark because it announced a special, rigorous standard for the admission of scientific evidence. Frye excluded the systolic blood pressure evidence because the theory had not gained "general acceptance in the particular field in which it belongs." Under Frye, it is not enough that the prosecutor can find one qualified expert to vouch for the new scientific technique. The prosecutor has to meet a much more stringent test; the prosecutor must establish that as a matter of historical fact, the theory has gained general acceptance within the relevant scientific circle.

During the 1970's, Frye proved to be a formidable barrier to the introduction of prosecution scientific evidence. For example, in a single year, 1977, the courts invoked Frye as the basis for excluding evidence of ion microprobe analysis, trace metal detection technique, and Decatur Ragun. The courts also relied upon Frye as the ground for banning evidence of forward looking infrared and hypnotic memory enhancement. In short, the decade of the 1970's was dominated by the threshold question of the admissibility of scientific evidence. Defense counsel invoked Frye so successfully that the commentators almost unanimously deplored the fact that Frye banned many promising forensic techniques from the courtroom.

The thesis of this article is that we are now entering a new stage in the evolution of scientific evidence—a stage that will be dominated by questions of the weight of scientific evidence rather than

19. Id.
20. Giannelli, supra note 1, at 1204-08.
22. Giannelli, supra note 1, at 1204-08.
23. Id.
27. United States v. Kilgus, 571 F.2d 508 (9th Cir. 1978).
admissibility. Part I of this article collects the data pointing to that conclusion. Part I notes that there is an emerging trend toward the liberalization of the standards for admitting scientific evidence. Part II of this article comes to grips with the import of that trend. If the courts are going to admit scientific evidence more liberally, we must learn to critically evaluate the weight of scientific evidence. Part II proposes an analytic approach to the weight of scientific proof.

I. A NEW ERA IN THE EVOLUTION OF SCIENTIFIC EVIDENCE—THE NECESSITY FOR LEARNING HOW TO EVALUATE THE WEIGHT OF SCIENTIFIC EVIDENCE

This modern era in the evolution of scientific evidence is paradoxical. Two seemingly inconsistent developments are occurring simultaneously. On the one hand, as previously noted, several jurisdictions have begun to relax the admissibility standards for scientific evidence. On the other hand, the relaxation is taking place at the very time that we are coming to a new, alarming realization of the incidence of error in forensic analysis.

The apparent contradiction is that this realization of the possibility of error in forensic analysis is dawning on us at the very time that many jurisdictions are beginning to liberalize the standards for admitting forensic proof. As we previously stated, Frye is the leading American authority on the admissibility of scientific evidence.30 Until very recently, Frye was the almost universal view among American courts. The overwhelming weight of authority in both federal and state courts accepted Frye,31 and it seemed to be the well-settled standard in at least forty-five states.32

Although the courts seemed satisfied with the prevailing Frye test, the commentators have been less than enthusiastic about Frye. Indeed, almost from its inception, Frye has been subjected to a drumbeat of criticism. First, the commentators claim that the courts have been unpredictable and selective in characterizing

30. Giannelli, supra note 1, at 1204.
32. Id.
proof as "scientific" evidence that must pass muster under Frye. Next, they charge that it is often difficult to identify the relevant scientific field. Should forensic science itself be treated as a scientific field? Third, the critics allege that the test's use of the expression, "general acceptance," renders the test ambiguous. The courts have never succeeded in quantifying the standard for general acceptance. Finally and most importantly, the commentators damn Frye because the test results in the exclusion of much valuable, reliable scientific evidence. The test ensures that the courts will constantly lag behind the advances of science while the courts wait for novel scientific techniques to win "general acceptance." By banning these techniques from the courtroom, Frye frustrates the search for truth.

These criticisms are having a telling impact on Frye. Notwithstanding the growing evidence of a substantial margin of error in forensic analysis, a number of jurisdictions have recently repudiated Frye and liberalized the standards for admitting scientific evidence. These jurisdictions have rationalized the liberalization on three theories.

Frye, of course, is a case law rule and has a common law origin. Some jurisdictions have abandoned Frye as a rule of decisional law. The first slippage away from Frye occurred in 1969 in Coppolino v. State, a Florida decision. The Florida court stressed that the trial judge has wide discretion in deciding whether to admit novel scientific evidence. During the 1970's, other jurisdictions largely ignored Coppolino. A full decade passed before another court, a New York Supreme Court, was daring enough to

33. Giannelli, supra note 1, at 1219-21.
34. Id. at 1208-10.
36. Giannelli, supra note 1, at 1215-19.
37. Id. at 1216.
38. Id. at 1223-24.
42. Id. at 70, 75; Giannelli, supra note 1, at 1234-35 ("Coppolino thus ignores rather than rejects Frye.").
overrule Frye. In mid-1980, in a case of first impression involving Human Leukocyte Antigen (HLA) blood test evidence, the Utah Supreme Court indicated that it would no longer treat general acceptance as an absolute requirement for the introduction of scientific evidence. The court reduced general acceptance to the status of a factor in a discretionary determination of whether the new forensic technique is trustworthy. In late 1980, the Iowa Supreme Court joined the ranks of the courts rejecting Frye. The court echoed all the criticisms of Frye and announced that “we do not believe that ‘general scientific acceptance’ is a prerequisite to the admission of [scientific] evidence . . . if the reliability of the evidence is otherwise established.”

While some courts have used their common law powers to jettison Frye, other jurisdictions have taken a second, statutory tack. On July 1, 1975, the new Federal Rules of Evidence became effective. Twenty-two jurisdictions have already adopted a version of the Rules, and more may follow suit soon. It is arguable that the Rules impliedly abolish Frye. Federal Rule 402 provides:

All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules or by other rules prescribed by the Supreme Court pursuant to statutory authority.

This provision precludes trial judges from inventing new, decisional rules excluding relevant evidence. Frye is a decisional rule, and the Rules nowhere codify Frye. Hence, there is a strong statutory construction argument that the Rules impliedly overrule Frye. Three jurisdictions—New Mexico, Maine, and the Court

43. People v. Daniels, 102 Misc. 2d 540, 422 N.Y.S.2d 832 (Sup. Ct. 1979).
45. Id. at 1234.
46. State v. Hall, 297 N.W.2d 80 (Iowa 1980).
47. Id. at 85.
49. Giannelli, supra note 1, at 1228-29 & n.241.
50. Id.
51. FED. R. EVID. 402.
52. Giannelli, supra note 1, at 1228-31. However, it should be noted that the legislative history of the Rules does not even faintly suggest that either Congress or the Advisory Committee intended to overrule Frye. 40 OHIO ST. L.J. 757, 763 n.53 (1979). Further, since the
of Appeals for the Second Circuit—have drawn that conclusion.

Lastly, other courts have invalidated Frye on constitutional grounds. In two cases, a straightforward due process argument met with success. In 1975, a New Mexico court accorded the defendant a due process right to introduce polygraph evidence, and in 1979 a federal district court reached the same result. There is significant scholarly sentiment favoring this due process rationale. There is another constitutional theory for admitting scientific evidence: the compulsory process guarantee. In Washington v. Texas and Chambers v. Mississippi, the Supreme Court found an implied right to present critical, reliable defense evidence in the express compulsory process provision. It was perhaps expectable that some court would eventually employ this sixth amendment rationale to override Frye and reverse a trial judge's exclusion of defense polygraph evidence. An Ohio trial court did precisely that. This decision gave a new constitutional impetus to the decline of the Frye test.

Concurrent with the decline of the Frye test was the growth of the realization that scientific proof is far from infallible. In the early 1970's, there were revelations of significant margins of error in toxicological analysis. There were indications, for example, of a substantial degree of error in blood alcohol analysis by both

adoption of the Federal Rules of Evidence, most of the courts of appeals that have addressed the issue have continued to apply Frye. 64 CORNELL L. REV. 875, 878-79 (1979).


55. United States v. Williams, 583 F.2d 1194 (2d Cir. 1978); see 40 OHIO ST. L.J. 757, 769 (1979).


60. Id.


clinical and public health laboratories. In the mid-1970's, Dinovo and Gottschalk reported significant interlaboratory variation in the quality of drug analysis.

These studies prompted the Law Enforcement Assistance Administration to sponsor a much broader and more systematic test of the proficiency of American crime laboratories. The LEAA's Laboratory Proficiency Testing Program began in fall 1974. Between 235 and 240 crime laboratories from throughout the United States participated in the program. To gauge the quality of analysis by these laboratories, a Project Advisory Committee prepared samples and sent them to the participating laboratories for analysis. The Committee knew the findings that a competent scientific analysis of the samples would yield. The Committee then judged the participating laboratories' reports against the known data.

It is an understatement to say that the findings of the Test Program are alarming. "Shocking" would be more precise. Below is a columnar analysis of the various tests and the percentage of "unacceptable" (inaccurate or incomplete responses).

On some of these tests, fewer than half the laboratories reached the correct conclusion.

---

65. Dinovo & Gottschalk, Results of a Nine-Laboratory Survey of Forensic Toxicology Proficiency, 22 CLINICAL CHEMISTRY 843-46 (1976).
66. Project Advisory Committee, Laboratory Proficiency Testing Program, Supplementary Report—samples 1-5, 6-10 at 1 (1975-76).
68. Law Enforcement Assistance Administration Newsletter (Sept. 1978).
69. See chart accompanying footnote 68 supra.
The LEAA project director, Mr. John O. Sullivan, declared that the Testing Program demonstrates that "[t]he laboratories are having difficulties in identifying the samples." The program unquestionably documented "a very real possibility of error in the forensic analyses conducted by police laboratories in the United States."  

This first development, the relaxed admissibility standards for

---

70. Law Enforcement Assistance Administration Newsletter 5 (Sept. 1978).  
71. Imwinkelried, supra note 3, at 637.
scientific evidence, makes it probable that defense counsel will be confronting more and more scientific evidence in the future. The pace of technological advance certainly has not slowed; and the common-law, statutory, and constitutional theories for overriding Frye should accelerate the rate at which forensic advances surface in the courtroom. Frye will no longer be the substantial barrier it was during the mid-1970's, and the opponent of the scientific evidence will have to resort more frequently to weight attacks.

The second development, the startling revelation of the incidence of error in forensic analysis, suggests that if properly prepared and presented, weight attacks can be successful. The trial judge may admit the proponent's scientific evidence, but the opponent may be able to persuade the trier of fact to attach little weight to the evidence. If the error levels disclosed in the Laboratory Proficiency Testing Program are at all typical of the error levels in most crime laboratories, weight attacks on the scientific evidence hold great promise.

To put the matter simply, the first development ensures that the opponent of scientific evidence will have more opportunities to attack the weight of the evidence; and the second development supplies the motivation to mount the weight attacks. The 1980's will witness a new epoch in the evolution of scientific evidence, and the hallmark of the new epoch will be a shift in focus from admissibility attacks to weight attacks.

As a case in point, consider the recent history of HLA evidence. A few years ago most criminal practitioners had not even heard of HLA evidence; the technique did not become widely known until the late 1970's. It was immediately apparent that HLA blood evidence has greater discriminatory power than the more traditional red cell systems such as ABO. The courts were so impressed with HLA that they readily accepted the evidence

74. Id. at 1137-43.
75. Id.
76. Seven states have held HLA test results admissible: California: e.g., County of Fresno v. Superior Court, 92 Cal. App. 3d 133, 154 Cal. Rptr. 660 (1979); Cramer v. Morrison, 88 Cal. App. 3d 873, 153 Cal. Rptr. 865 (1979); Florida: Carlyon v. Weeks, 387 So. 2d 465 (Fla. 1980).
and brushed aside Frye objections to HLA. The admissibility of HLA evidence was established very rapidly, and the struggle over HLA evidence quickly shifted from admissibility analysis to weight


Moreover, four states recently amended their blood test statutes to authorize the admission of HLA evidence. Georgia Code Ann. § 74-306 (1981) provides:

As soon as practicable after an action has been brought the court upon motion of the plaintiff, the defendant, or any other interested party, may order the mother, the alleged father, and the child to submit to any blood tests, including human leukocyte antigen (HLA) testing if available, which have been developed or established for purposes of disproving or proving parentage and which are reasonably accessible. If the court orders such blood tests and if the action is brought prior to the birth of the child, the court shall order the blood tests made as soon as medically feasible after the birth. The tests shall be performed by a duly qualified licensed practicing physician, duly qualified immunologist, or other qualified person. The court may, upon motion by a party, order that independent tests be performed by other experts qualified as examiners of blood types. In all cases, however, the court shall determine the number and qualifications of the experts. An order issued under this subsection is enforceable by contempt; except that if the petitioner refuses to submit to an order for a blood test, the court upon motion of the defendant may dismiss the suit. (emphasis added)

Indiana Code Ann. § 31-6-6.1-8 (Burns 1980) provides:

Upon the motion of any party, the court shall order all of the parties to the action to undergo either a blood grouping test or a Human Leukocyte Antigen (HLA) tissue test. The tests shall be performed by a qualified expert approved by the court, and the results of the tests may be received in evidence. (emphasis added)

Wisconsin Stat. § 767.48 was also amended to authorize generally the admission of evidence of "genetic markers," and the New York Family Court Act was reworded to sanction the introduction of "the results of the human leucocyte antigen blood tissue test." See In re Jane L., 7 Fam. L. Rep. (BNA) 2474 (N.Y. Fam. Ct., N.Y. County May 12, 1981).

78. Id.
The burning dispute over HLA evidence is now the extent to which the proponent will be permitted to quantify the weight of the evidence by citing a percent probability of paternity to the jury. The court's treatment of HLA augurs the way in which the courts will probably deal with novel scientific evidence in the 1980's: If the reliability of the evidence is demonstrable, the courts will waste little time in ruling the evidence admissible; and the real debate over the evidence in this decade will be the resolution of the probative weight of the evidence.

II. AN APPROACH TO EVALUATING THE WEIGHT OF SCIENTIFIC EVIDENCE

In the past, attorneys have been rather reluctant to attack the weight of scientific evidence offered against them. Until very recently, in drug cases most defense attorneys routinely stipulated to the results of the police laboratory analysis of the suspected drugs. At first, defense attorneys were also hesitant to attack sound spectrography evidence. One court noted that in eighty percent of the previous cases admitting sound spectrography evidence, the defense had not presented any expert testimony disputing the reliability of the voiceprint technique.

This hesitancy is understandable. The heavy emphasis on threshold admissibility questions in the reported cases certainly contributed to the neglect of weight issues. Moreover, most attorneys receive little or no law school training in evaluating the weight of evidence, including scientific evidence. Most evidence courses and texts concentrate on the admissibility of evidence, its legal sufficiency to sustain various burdens, and substitutes for evidence. The courses and texts completely overlook weight analysis. Trial techniques courses and texts are almost as neglectful. In

---

79. Ellman & Kaye, supra note 73, at 1143-61.
80. Id.
83. See, e.g., E. IMWINKELRIED, P. GIANNELLI, F. GILLIGAN & F. LEDERER, CRIMINAL EVIDENCE chs. 1-26 (1979) [hereinafter cited as CRIMINAL EVIDENCE]; C. McCORMICK, supra note 29, chs. 1-34.
84. CRIMINAL EVIDENCE, supra note 83, ch. 27; C. McCORMICK, supra note 29, ch. 36.
85. CRIMINAL EVIDENCE, supra note 83, ch. 28; C. McCORMICK, supra note 29, ch. 35.
large part, the typical trial techniques text simply teaches the evidentiary admissibility standards in a more practical fashion. The text will devote some coverage to attacking the weight of opposing evidence, but the evidence in question is usually the opposition's lay testimony.

Moreover, attorneys tend to feel more comfortable working with the Frye admissibility test than they do evaluating the weight of scientific evidence. Frye is an essentially historical test: As a matter of historical fact, has this forensic technique gained general acceptance within the relevant scientific circle? Attorneys constantly litigate historical issues, and they feel at ease litigating the question presented by the Frye test precisely because it is an historical issue rather than a scientific issue. An attorney can successfully litigate the application of the Frye test to a scientific technique even if the attorney knows very little about the technical aspects of the technique. However, when the focus shifts from admissibility to weight, the attorney needs a much more sophisticated understanding of the forensic technique.

Most of the published texts on scientific evidence reinforce the attorney's reluctance to evaluate the weight of scientific proof. The typical text is a catalogue of scientific techniques; the text lists one technique after another and presents a good deal of technical detail about each technique. These texts do not even attempt to synthesize the voluminous material and do not offer the attorney an overall approach to evaluating the probative weight of the myriad techniques mentioned in the texts. The attorney naturally is intimidated by the bewildering array of scientific techniques and data.

The purpose of Part II of this article is to outline an approach to the evaluation of the weight of scientific evidence. This approach groups forensic techniques into three broad categories: (1) instrumental techniques that yield numerical test results; (2) techniques that yield non-numerical, visual displays; and (3) "software" tech-

87. Id. at chs. III & V. Chapter V, devoted to expert witnesses, deals primarily with the admissibility of expert testimony.
niques. We shall now review each category, list some of the techniques falling within the category, and identify the pivotal questions that must be asked to properly evaluate the probative weight of the technique.

A. The First Category: Instrumental Techniques that Yield a Numerical Result

In the minds of most lay jurors, these techniques epitomize scientific evidence. Laypersons ordinarily assume that truly precise—and, hence, “scientific”—test results can be couched in numerical terms. They are most impressed by sophisticated instrumentation capable of exact, minute measurement. There are numerous illustrations of this type of technique. One example is the gas chromatograph (GCI) used to detect tiny quantities of alcohol vapor in breath samples.\textsuperscript{89} Another illustrative technique is neutron activation analysis (NAA).\textsuperscript{90} One of the primary advantages of this technique is its tremendous sensitivity.\textsuperscript{91} NAA can measure quantities 1,000 times smaller than prior techniques.\textsuperscript{92} Although its sensitivity varies with the element being analyzed, the chart below shows NAA’s fantastic capability for measuring minute samples:\textsuperscript{93}

\begin{quote}
\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
Element & Sensitivity (ppm) & Prior Technique & NAA
\hline
Metal & 1,000 & \ldots & \ldots
\hline
Other & \ldots & \ldots & \ldots
\hline
\end{tabular}
\end{center}
\end{quote}

\textsuperscript{90} A. Moenssens & F. Inbau, \textit{supra} note 88, ch. 9.
\textsuperscript{92} \textit{Id.} at 124.
\textsuperscript{93} \textit{Id.}
One hour irradiation at a thermal neutron flux of
\[ 2 \times 10^{13} \text{ n/cm}^2 \text{ sec} \]

<table>
<thead>
<tr>
<th>Element</th>
<th>Limit of Measurement (micrograms)</th>
<th>Element</th>
<th>Limit of Measurement (micrograms)</th>
<th>Element</th>
<th>Limit of Measurement (micrograms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>0.01</td>
<td>Indium</td>
<td>0.00005</td>
<td>Rhodium</td>
<td>0.0005</td>
</tr>
<tr>
<td>Antimony</td>
<td>0.005</td>
<td>Iodine</td>
<td>0.005</td>
<td>Rubidium</td>
<td>0.05</td>
</tr>
<tr>
<td>Argon</td>
<td>0.05</td>
<td>Iridium</td>
<td>0.0001</td>
<td>Ruthenium</td>
<td>0.01</td>
</tr>
<tr>
<td>Arsenic</td>
<td>0.001</td>
<td>Iron</td>
<td>200</td>
<td>Samarium</td>
<td>0.0005</td>
</tr>
<tr>
<td>Barium</td>
<td>0.05</td>
<td>Lanthanum</td>
<td>0.001</td>
<td>Scandium</td>
<td>0.01</td>
</tr>
<tr>
<td>Bismuth</td>
<td>0.5</td>
<td>Lead</td>
<td>10</td>
<td>Selenium</td>
<td>0.01</td>
</tr>
<tr>
<td>Bromine</td>
<td>0.005</td>
<td>Lithium</td>
<td>100</td>
<td>Silicon</td>
<td>0.05</td>
</tr>
<tr>
<td>Cadmium</td>
<td>0.05</td>
<td>Lutecium</td>
<td>0.00005</td>
<td>Silver</td>
<td>0.0001</td>
</tr>
<tr>
<td>Calcium</td>
<td>1</td>
<td>Magnesium</td>
<td>0.5</td>
<td>Sodium</td>
<td>0.005</td>
</tr>
<tr>
<td>Cerium</td>
<td>0.1</td>
<td>Manganese</td>
<td>0.00005</td>
<td>Strontium</td>
<td>0.005</td>
</tr>
<tr>
<td>Cesium</td>
<td>0.5</td>
<td>Mercury</td>
<td>0.01</td>
<td>Tantalum</td>
<td>0.05</td>
</tr>
<tr>
<td>Chlorine</td>
<td>0.01</td>
<td>Molybdenum</td>
<td>0.1</td>
<td>Tellurium</td>
<td>0.05</td>
</tr>
<tr>
<td>Chromium</td>
<td>1</td>
<td>Neodymium</td>
<td>0.1</td>
<td>Terbium</td>
<td>0.05</td>
</tr>
<tr>
<td>Cobalt</td>
<td>0.005</td>
<td>Nickel</td>
<td>0.05</td>
<td>Thallium</td>
<td>5</td>
</tr>
<tr>
<td>Copper</td>
<td>0.001</td>
<td>Niobium</td>
<td>0.005</td>
<td>Thorium</td>
<td>0.05</td>
</tr>
<tr>
<td>Dysprosium</td>
<td>0.0000001</td>
<td>Osmium</td>
<td>0.05</td>
<td>Thulium</td>
<td>0.01</td>
</tr>
<tr>
<td>Erbium</td>
<td>0.001</td>
<td>Palladium</td>
<td>0.05</td>
<td>Tin</td>
<td>0.5</td>
</tr>
<tr>
<td>Europium</td>
<td>0.000005</td>
<td>Phosphorus</td>
<td>0.5</td>
<td>Titanium</td>
<td>0.05</td>
</tr>
<tr>
<td>Fluorine</td>
<td>1</td>
<td>Platinum</td>
<td>0.05</td>
<td>Tungsten</td>
<td>0.001</td>
</tr>
<tr>
<td>Gadolinium</td>
<td>0.01</td>
<td>Potassium</td>
<td>0.05</td>
<td>Uranium</td>
<td>0.005</td>
</tr>
<tr>
<td>Gallium</td>
<td>0.005</td>
<td>Praseodymium</td>
<td>0.00005</td>
<td>Vanadium</td>
<td>0.001</td>
</tr>
<tr>
<td>Germanium</td>
<td>0.005</td>
<td>Rhenium</td>
<td>0.00005</td>
<td>Ytterbium</td>
<td>0.001</td>
</tr>
<tr>
<td>Gold</td>
<td>0.0005</td>
<td></td>
<td></td>
<td>Zinc</td>
<td>0.1</td>
</tr>
<tr>
<td>Hafnium</td>
<td>1</td>
<td></td>
<td></td>
<td>Zirconium</td>
<td>1</td>
</tr>
<tr>
<td>Holmium</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Activation Analysis Sensitivities. The detection limits of many elements listed here can actually be pushed 10 to 100 times further if needed.

This is obviously the sort of evidence that may awe the trier of fact. Consider a simple hypothetical. An NAA expert takes the stand and first describes the fantastic capabilities of the instrument. The expert then testifies that she ran a comparative analysis of the paint on defendant's car and a paint chip found at the crime scene. The analysis revealed the "same" minute quantities of three different chemical elements in the two samples. How can we evaluate this evidence? How can we prevent the jury from leaping to the conclusion that the paint chip necessarily came from the defendant's car? To properly evaluate this type of evidence, we must ask two questions.
1. Are there any imprecisions in the formula the instrument uses to compute the number?

There is such an imprecision in the case of the gas chromatograph. The GCI directly measures breath alcohol and then employs a formula to convert the breath alcohol figure into a blood alcohol concentration (BAC).\textsuperscript{94} The GCI makes the conversion on the assumption that there is the same amount of alcohol in one part of the subject's blood as there is in 2,100 parts of the subject's breath.\textsuperscript{95} Is the 2,100 coefficient universally true? The answer is no. Research indicates that the blood-breath coefficient varies from person to person.\textsuperscript{96} The lower limit seems to be a coefficient of 1,500/1, and the upper limit seems to be a 3,000/1 coefficient.\textsuperscript{97} Hence, the number generated by the GCI is not entitled to conclusive weight on the question of the subject's blood alcohol concentration.\textsuperscript{98}

There are other imprecisions in the formulae used in blood alcohol analysis. Assume \textit{arguendo} that the GCI yielded the correct BAC at the time of the GCI test. The test is usually administered some time after the legally relevant event, that is, the assault or traffic accident. To compute the BAC at the time of the legally relevant event, the analyst must work backwards in time. To make the retrospective computation, the analyst normally assumes that the subject's clearance or elimination rate is 0.015\% per hour.\textsuperscript{99} Is that percent a constant? Again, the answer is no. The available data indicates that the subject's clearance rate can range from 0.006\% to 0.04\%.\textsuperscript{100} Even if the BAC reading itself were absolutely correct, the retrospective BAC is not entitled to determinative weight.

\textsuperscript{94} Hutton, \textit{supra} note 89, at 221.
\textsuperscript{95} \textit{Id}.
\textsuperscript{96} Imwinkelried, \textit{Scientific Evidence Update}, in SIXTEENTH ANNUAL DEFENDING CRIMINAL CASES 707, 752 (citing research by Doctors Dubowski and Mason).
\textsuperscript{97} \textit{Id}.
\textsuperscript{98} \textit{Id}.
\textsuperscript{100} \textit{Id} at 284.
2. Is there an adequate comparative data base to properly evaluate the statistical significance of the number?

NAA does not suffer from the same weaknesses as the GCI. Unless there is interference masking one of the elements in the sample, NAA will give a true reading of the precise quantities of the various elements present; the numbers will represent the correct elemental composition of the sample being analyzed. But does that mean that if the analyst finds the same quantities of elements in the two paint samples, we must infer that the samples came from the same source? Or if NAA reveals the presence of antimony and barium, the components of gunshot residue, must we conclude that the subject recently fired a weapon? Common sense and sound statistical analysis supply the same answer: no.

In the past, some NAA experts have attempted to express such conclusive opinions to lay jurors. But, upon reflection, it is clear that in framing those opinions, the analysts were guilty of classic non sequiturs. We cannot draw those inferences until we have developed a comparative data base. It may be true that there are antimony and barium on the subject's hand and that they are the components of primer. However, we must not leap to the conclusion that the subject recently fired a weapon. First, we must ask: In what quantities do antimony and barium naturally occur in the environment? What quantities of antimony and barium might an innocent person collect on his or her hands through normal environmental or occupational exposure? If the quantity of antimony detected on the subject's hands is equal to or lower than the "handblank" value, that is, the quantity he or she might pick up through innocent exposure, the NAA finding does not dictate the conclusion that the subject recently fired a weapon.

The only way to determine whether we should draw the inference is to empirically investigate the environmental and occupa-

101. Imwinkelried, supra note 96, at 759-60.
106. Id.
tional incidence of those elements. In response to statisticians' critique of past NAA testimony, several law enforcement and research laboratories undertook studies of handblank values. These studies now enable us to draw much more confident inferences from NAA of suspected gunshot residue. However, there are other areas in which the data base is still undeveloped. For example, there has been relatively little research on the application of NAA to paint samples. The studies that have been conducted have involved relatively small data bases, and the studies themselves caution against drawing inferences from their limited research. When an extensive data base is available, the NAA data is entitled to considerable weight; but absent such a base, we should be wary of attributing too much significance to the bare numbers.

B. The Second Category: Techniques that Produce Visual Displays

The techniques in the first category yield one type of visual display: a digital readout. There are other techniques that do not yield numbers but nevertheless produce some sort of visual display. Among the most common examples are chemical color tests, polygraphy, and sound spectrography. For example, the most common sort of drug identification test is a chemical color test. In this test, the analyst adds a drop of known reagent to some of the unknown drug, and the reaction produces a color—a color that may be indicative of the identity of the unknown. Similarly, in

107. Id.; Cowan & Purdon, supra note 104; Imwinkelried, supra note 96, at 760-61.
110. Id.
111. In Snow & Washington, supra note 109, the authors assert that “due to the limited sample population, the results of this study cannot necessarily be extrapolated to all paints, and probabilities of elemental occurrence cannot be assuredly assigned.” Id. at 919.
113. Id.
polygraphy, the end product is a polygram chart with numerous markings to depict the subject's physiological reactions during the questioning.\textsuperscript{114} And in sound spectrography, the technique generates a spectrogram, which is a visual display of time and the voice's pitch and amplitude.\textsuperscript{115} Precisely because the display is non-numerical, the interpretation of the display is a more subjective and difficult task than interpreting the readout from an instrument in the first category. Reading the number is a purely mechanical task while evaluating a polygram or spectrogram requires more subtle interpretation. To determine what weight to attach to the analyst's evaluation, we must examine the validity and reliability of the analyst's interpretive standards.

Although laypersons and attorneys often use the expressions, "validity" and "reliability," interchangeably, validity and reliability are distinct concepts in forensic science.\textsuperscript{116} "Validity" refers to the technique's ability to measure what it is supposed to measure; a technique's validity is its accuracy.\textsuperscript{117} Sound spectrography is designed to enable the analyst to determine whether the same voice produced two spectrograms. The technique's validity depends upon the percentage of cases in which the analyst correctly makes the determination. In contrast, "reliability" refers to the consistency of the technique.\textsuperscript{118} Thus, a technique's reliability depends upon the percentage of cases in which independent examiners will draw the same inference from the test result. If two equally competent polygraphists, one in New York and the other in California, examine the same polygrams, how often will their diagnoses of deception agree?\textsuperscript{119} To properly evaluate this category of forensic technique, we must learn to assess the validity and reliability of the analyst's interpretive standards.

1. \textit{How valid is the analyst's interpretive standard?}

A technique is invalid to the extent that it produces false posi-
tives or false negatives. A false positive is an incorrect identification. Suppose that two different speakers produced the two spectrograms. The analyst’s opinion would be a false positive if the analyst concluded that the same person produced the spectrograms. A false negative is the converse: It is an incorrect elimination. Now assume that the same speaker produced both spectrograms. The analyst’s opinion would be a false negative if the analyst concluded that two different voices produced the spectrograms.

The validity of a forensic technique can be assessed by quantifying the expectable percentage of false positives and false negatives. After doing so, we can determine the weight to be assigned to the evidence because we know the risk of error. Thanks in large part to the research of Dr. Oscar Tosi, we can define the risk of error for at least some applications of the sound spectrography technique. Tosi conducted well-designed experiments to ascertain the percentage of false positives and negatives in sound spectrography. His research produced the finding that on the average, the false positive figure was six percent, but that if the examiner is permitted to express no opinion when he or she thinks that the question is too close, the false positive figure falls to two percent. Tosi also concluded that on the average, the false negative figure was thirteen percent, but that when the examiner had the option of venturing no opinion, the false negative figure dropped to five percent. At this point, the question whether those error rates are too high to admit voiceprint is irrelevant; we are focusing now on weight rather than admissibility. The point is that Tosi’s
research enables us to appraise the validity and deserved weight of sound spectrography evidence.

Suppose, on the other hand, that we know that a forensic technique can yield false positives and negatives, but the available research data does not permit us to quantify the margin of error. For instance, we know that many of the chemical color tests for drug identity yield false positives,\textsuperscript{128} but the percentage of false positives has not been quantified with any degree of precision. In this case, little weight should be attached to this evidence; we know that the technique is prone to error, but we have no objective method of estimating the risk of error.

2. How reliable is the analyst's interpretive standard?

The reliability of the interpretive standard for evaluating this type of forensic evidence is often the key to deciding the weight of the evidence. As previously noted, the non-numerical nature of the test result injects an inexorable element of subjectivity into the interpretation of the test. The test's reliability determines the extent to which the test is flawed by subjectivity.\textsuperscript{129} In the technical sense of the term, "reliability" is the antithesis of subjectivity. If independent analysts separated by thousands of miles can look at the same visual display and draw the same inferences from the display, the test's subjectivity is minimal; and the non-numerical character of the test result should not deter us from attaching great weight to the evidence.

The criterion of reliability distinguishes accepted techniques such as fingerprint analysis from such maligned forensic techniques as polygraphy. Fingerprint identification ranks high in reliability. There exists a detailed system for the primary and secondary classification of fingerprints,\textsuperscript{130} and this classification system helps the analyst isolate distinctive points of similarity between two fingerprints. In the United States, there is no set number of


\textsuperscript{129} Abrams, \textit{supra} note 116, at 791-99; Giannelli, \textit{supra} note 1, at 1201 n.20.

points of similarity required for an identification, but ten points is the standard most commonly preferred. Hence, there is substantial consensus among fingerprint analysts on what constitutes a distinctive point and how many points are needed for identification. This consensus has made fingerprint identification a relatively reliable technique, and fingerprint analysts rarely disagree. Equally competent fingerprint analysts working independently in Virginia and Missouri should duplicate each other’s results and reach the same conclusion.

The likelihood of disagreement is much greater in the case of polygraphy. As one experienced polygraphist has candidly conceded, “The [polygraph] technique is not like fingerprints . . . where experts miles apart can come to the same conclusions in examining the same set of prints . . . .” Research data indicate that polygraphists’ interpretive standards are less reliable and more subjective. Even the leading proponents of polygraph evidence admit that there have been fewer reliability studies of polygraphy than validity studies. They also concede that some of the leading reliability studies show that in as many as fifteen percent of the cases, polygraphists are likely to disagree over the interpretation of the polygram. Polygraph critics point to a recent Ohio case as a perfect illustration of the unreliability of polygraphy. In that case, four of the leading polygraph experts in the nation examined the polygrams and reached radically different conclusions.

If a technique’s interpretive standards have been experimentally verified as being reliable, the expert’s interpretation is entitled to

131. Id. at 654.
132. Id. at 652.
133. J. RICHARDSON, MODERN SCIENTIFIC EVIDENCE § 18.1 (1974) (“Scientific authority now emphatically declares that fingerprints are reliable as a means of identification.”); id. § 18.2 n.9 (“The fingerprints of an individual are personal to him, without duplication in the prints of any other person.”).
136. Id. at 798.
137. Id. at 798-99.
139. Id.
great weight. If that expert witness is qualified, the probability is that other competent experts would draw the same inferences. However, if the technique produces a non-numerical result and there is inadequate proof of the reliability of the interpretive standard, the expert’s interpretation should be regarded with caution. In this situation, we do not have the same assurance that other qualified experts would concur with the witness on the stand. When faced with the task of evaluating the weight of a scientific technique that yields a non-numerical result, the opponent should demand to know the extent to which the reliability of the technique’s interpretive standards has been established.

C. The Third Category: “Software” Techniques

Our third category includes what we shall label “software.” The label is admittedly imprecise but at least suggests the common denominator of the techniques falling in this category: No instrument intervenes to produce a numerical result such as a digital BAC readout or a visual display such as a polygram. “Software” includes such forensic findings as a psychiatrist’s diagnosis of mental illness or a pathologist’s characterization of a wound as homicidal. The analyst directly observes the primary data and formulates an opinion without resorting extensively to any intermediate instrument.

This is the sort of evidence least likely to awe the jury. When a layperson thinks of science, the layperson naturally thinks of sophisticated instruments capable of precise management. Software techniques are the farthest removed from the layperson’s conception of science; and for that reason, in the minds of many layperson these techniques hardly deserve the august title, “scientific.” The element of subjectivity in these techniques is patent to any juror. Yet that observation is small comfort to the attorney opposing software evidence; many types of software evidence are admissible in the courtroom,140 and the attorney will have great difficulty evaluating the weight of software evidence. As we did with the other two categories of forensic techniques, we shall identify the questions the attorney should ask to determine that weight.

140. See A. MOEnSSENS & F. INBAU, supra note 88, chs. 3 & 5. But see id. ch. 15.
1. How reliable is the analyst's interpretive standard?

As in the case of techniques producing non-numerical visual displays, one of the critical questions is the reliability of the analyst's interpretive standards. The psychiatric field highlights the importance of this question, for there has been much criticism of the unreliability of psychiatric interpretive standards. Some of the criticism has come from judicial circles. Both Justice Burger and Judge Bazelon have pointed to the embarrassing frequency with which psychiatrists disagree in the courtroom. This criticism finds support in the psychiatric and psychological literature. That literature contains numerous studies indicating that the political and social biases of the psychiatrist can greatly influence the psychiatrist's evaluation. In one recent research project, the psychiatrists at a clinic were found to agree on specific diagnoses in only twenty-one percent of the cases. They disagreed in thirty-one percent of the cases. Given the state of the record, psychiatric testimony should generally be accorded little weight. The reliability of most psychiatric opinion is of a much lower caliber than the proven reliability of, for example, fingerprint identification.

Yet we cannot generalize about all psychiatric evaluation. We must focus on the precise technique the psychiatrist is using; and if we do, we shall occasionally find that there are relatively objective standards for evaluating the outcome in some psychiatric techniques. One frequently used personal assessment technique is the Minnesota Multiphasic Personality Inventory (MMPI). The test consists of 566 statements which the subject marks true or false. However, in interpreting the test results, the psychiatrist does not simply make an intuitive, impressionistic judgment. There are

144. See the numerous articles cited in Ziskin, supra note 141.
145. Id.
146. Shell, Psychiatric Testimony: Science or Fortune-Telling?, BARRISTER, Fall 1980, at 8.
147. Id.
148. Ziskin, supra note 141, at 1108-09.
scales for grading the test results. On the basis of the test answers, the subject is compared with a normative population and assigned a position on each scale. The positions on the scales correspond roughly with various psychiatric diagnoses. Numerous studies have empirically established correlations between certain configurations of scale scores and particular psychiatric diagnoses. These scales remove much of the subjectivity from this psychiatric evaluation.

A concerted effort is being made to improve the reliability of interpretive standards in psychiatry. The American Psychiatric Association recently released *DSM III*, the third edition of its *Diagnostic and Statistical Manual of Mental Disorders*. The new edition dramatically revises the symptomatology for some of the leading psychiatric diagnoses. In the past, the *DSM* included vague lists of symptoms without requiring particular symptoms or a specific number of symptoms to support a diagnosis. The vague provisions in *DSM* understandably resulted in unpredictable and unreliable diagnoses. Several researchers were troubled by this unreliability and worked to make the symptomatology more objective. The end product of their work was the development of the so-called Feighner criteria, which have largely been incorporated in *DSM III*. These criteria often dictate that a certain number of listed symptoms must be present for a particular diagnosis. These criteria promise to increase the reliability of diagnosing these mental illnesses.

When the witness cannot point to hard proof of the reliability of his or her interpretive standard, the trier of fact should ascribe little weight to the expert's interpretation. Unfortunately, that is

---

149. Id.
150. Id.
151. Id.
152. Id.
154. Spitzer, Endicott & Robins, Clinical Criteria for Psychiatric Diagnosis and *DSM-III*, 132 AM. J. PSYCH. 1187 (1975); Feighner et al., Diagnostic Criteria for Use in Psychiatric Research, 26 ARCHIVES GENERAL PSYCH. 57 (1972).
155. Id.
156. Id.
often the case in the psychiatric field. However, as we have seen, rather than summarily dismissing all psychiatric evidence, the attorney should investigate to determine whether the particular evaluation the psychiatrist made is one of the exceptional determinations governed by relatively objective criteria. In a surprisingly large number of cases, the attorney will find that to be the case. When the mental health professional employs interpretive standards of demonstrated reliability, the interpretation is entitled to weight comparable to that accorded the most reliable evidence in the second category.

2. *Has the expert collected all the factual data necessary to have a reliable basis for the opinion?*

The softer the expert's interpretive standard, the more critical it becomes to assess the hardness of the factual basis of the opinion. Especially in the field of forensic psychiatry, the expert's evaluation must "be anchored in and tied to as much hard information as possible." As one of the leading texts on law and psychiatry emphasizes, the value and weight of a psychiatric evaluation are vitally dependent on the extent to which the expert rests the evaluation on concrete, historical data.

Although the experts in a field may disagree on the interpretive standard to apply to the underlying, hard data, they may agree on the type of data needed to form any reliable opinion. The field of forensic pathology is illustrative. Many of the determinations made by the forensic pathologist are software findings in the same sense as findings by a forensic psychiatrist. Indeed, one well known pathologist has described an autopsy as "an exercise in objective observation and subjective interpretation." For example, when the pathologist is asked to determine whether wounds on the cadaver were produced by a particular type of instrument or were homicidal rather than suicidal, there are no mechanical formulae

159. *Id.* at 1103.
the pathologist can use to make the determination. However, experienced practitioners would probably agree that the pathologist must gather certain types of data before forming any opinion in the case. For instance, many forensic pathologists would concur that a microscopic tissue examination is necessary before forming any interpretation of the wound pathology. Forming an opinion without that hard, supporting data would be "extremely dangerous." If the pathologist were sloppy or foolish enough to frame an opinion without gathering that data, the trier of fact should severely discount the opinion.

In short, the attorney analyzing software evidence should consider both the interpretive standard and the underlying data the standard is applied to. The attorney will sometimes find that even when experts disagree violently over the controlling interpretive standard, they are in virtually unanimous agreement over the types of hard data that are needed to have a trustworthy basis for any opinion. If an expert rushes to an opinion before collecting all the requisite data, the result should be the same as when the expert cannot demonstrate the reliability of the interpretive standard: The trier of fact should conclude that the expert's opinion deserves little credence.

III. CONCLUSION

Before closing this article, two caveats are in order. One is that the questions we have identified in Part II of this article are by no means the only questions relevant to the admissibility and weight of scientific evidence. The approach outlined in this article does not purport to be an exhaustive list of all the relevant factors. Many other factors, including the instrument's working condition, the witness' expertise, and the use of proper test procedures, bear upon the admissibility and weight of scientific evidence. However, those factors should be obvious even to attorneys with no scientific background. We have attempted to highlight the subtler

163. Id.; A. MOENSSSENS & F. INBAU, supra note 88, ch. 5.
165. Id.
166. CRIMINAL EVIDENCE, supra note 83, ch. 8.
questions that are not as readily apparent.

The second caveat is that our three categories are not necessarily mutually exclusive. For example, the results of some tests in the second category could easily be converted into numerical terms. Chemical color tests for drug identification fall into the second category, but we could convert the visual display into numerical terms by describing the resulting color in terms of wavelength and intensity.\(^{167}\) Similarly, the polygraph falls into the second category, but we could measure the periodicity and amplitude of the patterns on the polygram.\(^{168}\) We have classified forensic techniques on the basis of the manner in which the analyst usually conducts and interprets the test rather than on the basis of whether the test result could possibly be couched in numerical terms.

Notwithstanding these caveats, the approach outlined should prove useful to attorneys. In a pragmatic sense, the approach works. Experience suggests that in the case of most forensic techniques, this approach will help the attorney in identifying the key questions determining the weight of the evidence. For example, we placed neutron activation analysis (NAA) in the first category, in which one of the central questions is whether there is an adequate comparative data base. The most astute students of NAA have commented that the sufficiency of the comparative data base is the primary problem in modern NAA.\(^{169}\) By the same token, we assigned polygraphy to the second category, in which one of the major issues is the reliability of the interpretive standards. The most perceptive students of polygraphy have identified that very issue as the heart of the debate over polygraphy.\(^{170}\) Our categorization is far from perfect; but in many, if not most, instances, the categorization will lead the attorney to ask the right questions about the weight of the scientific evidence.

This approach is not simply an approach to attacking scientific evidence. The approach is certainly useful from the perspective of the opponent of the evidence; the approach should aid the oppo-

---

168. Id.
tent in identifying points of attack. However, more fundamentally, this is an approach to evaluating the weight of scientific evidence. If there are no imprecisions in the process an instrument uses to yield a number and we do have an extensive comparative data base, that evidence is entitled to great weight; and if the proponent clearly explains the precision of the process and the extent of the data base, the trier of fact will probably have the good sense to accord the evidence that weight.

Although this approach will be helpful to attorneys, it would be dishonest to suggest that this approach will make it easy for attorneys to evaluate the weight of scientific evidence. In the last era of scientific evidence dominated by admissibility questions, attorneys could be content to have a superficial understanding of forensic science. In most cases, the outcome turned on the application of Frye, and the attorneys litigated historical issues rather than properly scientific issues. But in the coming era, that will change. Scientific evidence will be admitted more readily, the focus will shift to the weight of the evidence, and attorneys will need a much more in-depth appreciation of forensic science to properly analyze the weight of the evidence. Juries will be exposed to scientific evidence more frequently; and in order to teach jurors the utility and limitations of forensic science, attorneys themselves will have to become more avid students of science.

This educational process is complicated by the fact that the process is ongoing at several levels. Forensic scientists themselves are devoting more energy to the problem of evaluating the weight of scientific evidence. As one leading forensic scientist recently pointed out, the 1970's witnessed the development of new technical tools to conduct forensic work "but as yet we have failed to define in statistical terms the meaning and significance of the data they generate."\textsuperscript{171} The challenge to the scientific community is to expand its comparative data bases and refine its interpretive standards.\textsuperscript{172} For their part, in this new era of scientific evidence attorneys must not only be familiar with the legal standards governing

\textsuperscript{171} Saferstein, Criminalistics—A Look Back at the 1970s, a Look Ahead at the 1980s, 24 J. For. Sci. 925, 930 (1979).

\textsuperscript{172} Organizations such as the Forensic Sciences Foundation and the Committee on Forensic Sciences of the American Society for Testing and Material are pursuing this objective.
the admissibility of scientific proof; attorneys must also familiarize themselves with the data bases and interpretive standards that facilitate evaluating the weight of scientific evidence. Only a cooperative, educational partnership between forensic experts and attorneys will ensure that the advent of this new era in scientific evidence will have a salutary effect on our legal system.