CHAPTER 1

Standards of Legal Admissibility and Their Implications for Psychological Science

David L. Faigman and John Monahan

Although jurisdictions employ a dizzying array of admissibility standards to assess the value of expert evidence, most of these are variants of three basic tests. The first originated in *Frye v. United States* (1923) and inquires whether the expert opinion is based on knowledge or a technique that is generally accepted in the particular field from which it comes. The second test, which has largely become predominant today, originated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), and inquires whether the expert testimony is based on methods and principles that are likely to produce valid opinions. The third test, used by sporadic jurisdictions and sporadically by *Frye* and *Daubert* jurisdictions as an exception to their usually applicable admissibility standards, is referred to as the "relevancy test." This test is not much of a test at all, however, since it inquires only whether the expert opinion is relevant and the witness is competent (i.e., qualified), conditions that are prerequisites for the admission of all evidence. In one form or another, and either alone or in combination, these three tests are used by virtually all U.S. jurisdictions today to measure psychological expert testimony.

It must be noted at the outset, however, that these tests represent archetypes and that few jurisdictions use a pure form of any one of them. Indeed, *Daubert* itself advanced the general acceptance criterion as one of several that

might be used to measure the validity of the underlying basis for proffered expertise. Moreover, some states use the *Frye* or *Daubert* test for scientifically derived expert testimony, but employ the relevancy test for expert opinion based on personal experience. In California and Arizona, for example, this exception exempts from serious threshold review psychological expert opinion that is deemed "experience-based." Scientifically grounded psychological expertise, in contrast, is evaluated under a rigorous threshold test in these states. In this chapter we consider the implications of these basic admissibility standards, in their pure and hybrid forms, for psychological science.

In the next section we begin by exploring the time before *Daubert,* when the "general acceptance" test of *Frye* provided the lodestar that guided the introduction of scientific evidence in court. The *Frye* test remains highly relevant today, though it has changed fundamentally due to the new understanding brought about by *Daubert*. Later, we turn more fully to the *Daubert* decision and subsequent developments in case law and in amendments to the Federal Rules of Evidence. *Daubert* brings fairly high expectations to expertise based on psychological science, and this section explores the implications of these demands on the psychology and law connection. Increasingly, however, courts exempt aspects of psychological expertise from review under admissibility tests. For example, states such as California and Arizona explicitly exempt experience-based expert opinion and apply the weak relevancy test to this testimony. We also consider the justification for this practice of exempting experience-based expert opinion and critically assess the wisdom of this practice. Finally, in conclusion, we consider the future prospects for psychology and psychologists under these admissibility standards.

FRYE'S GENERAL ACCEPTANCE TEST

All rules of admissibility are constructed on the premise that, although U.S. courts employ the adversarial process, the parties do not fully control what evidence reaches the trier of fact. In the realm of expert evidence, this means that some principle must control what sorts of evidence are admitted and what sorts are excluded. Prior to *Daubert*'s validity test, and still the most often posited alternative to that test, the general acceptance standard of *Frye* provided the basic procedure for handling scientific and technical evidence. Under this standard, novel expert testimony is admissible only once it has reached general acceptance in the particular field from which it comes. Although few courts cited *Frye* immediately after it was

decided, general acceptance eventually became the standard to which all subsequent tests of admissibility have been compared (Faigman, Porter, & Saks, 1994).

Courts'—and scholars'—views of *Frye* are decidedly mixed. *Frye* plainly possesses an assortment of positive attributes. Foremost, perhaps, the general acceptance test does not require judges and lawyers to have any substantive knowledge of the science underlying proffered expertise. Lawyers are not generally known for their proficiency with science (Gatowski et al., 2001). The *Frye* test puts little pressure on them to learn any. General acceptance employs the logical proposition that the best groups to ask about a technology or set of research findings are the very scientists who specialize in the subject. Waiting for a consensus to develop also operates to insulate courts from changing fads and fashions. The law is principally a conservative discipline and, on the whole, judges would prefer not to amend or overrule settled law with the publication of each new study or each new edited volume. *Frye* seemed to ensure an orderly and accurate, albeit gradual, development of the jurisprudence of scientific evidence.

But *Frye* also has more than its share of detractors. First of all, *Frye*'s perceived strength of not requiring judges and lawyers to know much about science also means that the law effectively abdicates review authority to others. *Frye* also can be readily manipulated in order to obtain preordained outcomes, and for that reason has been labeled both too conservative and too liberal. It is too conservative because it can lead courts to exclude cutting-edge research until the respective fields embrace the new work. True consensus can take years to develop, even for the best science. For example, despite the general acclaim among psychologists for eyewitness identification research (Kassin, Tubb, Hosch, & Memon, 2001), courts sometimes exclude expert opinion on this subject because it continues to attract significant detractors (*People v. Smith*, 2004). At the same time, however, *Frye i*s sometimes criticized for being too liberal, in that it might permit experts to testify to opinions based on little or no research. Because *Frye* only consults the pertinent field, findings from scientific disciplines that bring little rigor to their claims might pass legal muster quite readily despite the absence of adequate data or poor research methods. Although many examples might be cited, perhaps the most notorious is the battered woman syndrome. It has achieved widespread judicial acceptance based largely on a consensus of experts who are devoted to the subject, despite a research foundation that is very thin (Faigman, Kaye, Saks, Sanders, & Cheng, 2007). In this way, *Frye* effectively constitutes deference to established professional fields, and admissibility becomes a guild issue (Faigman, 1989).

Moreover, the *Frye* inquiry contains an inherent ambiguity. Judges are prompted to ask whether particular findings or techniques are generally accepted among researchers and practitioners in the field. But the relevance of the science may be very different for the law than it is for the researchers and practitioners who are asked about general acceptance. Consider, for example, polygraphs. Polygraph utility for the courts depends wholly on the test's power to distinguish between truthful answers and those that are deceptive for particular questions asked. Use in the field includes this utility, but may also include many others, including as a prop to interrogations, deterrence in national security screening, and detection of general nervousness to a line of questioning. Inquiring about the general acceptance of a technique without being specific regarding what it is generally accepted to do is a source of great confusion under *Frye*. There is little question that polygraphs are generally accepted among police, government security agencies, and professional polygraphers. But the acceptance of the polygraph as a tool of interrogation says little about its scientific validity (National Research Council, 2003). Psychology, perhaps more than most sciences, is rife with examples of this phenomenon. A diagnosis may be generally accepted for therapeutic purposes, for example, but not for forensic purposes. Rape trauma syndrome illustrates this therapeutic/forensic divide, since the underlying trauma associated with the syndrome is largely unquestioned in therapy but is the operative issue in the courtroom. Judges must ask the right questions under *Frye* in order to get useful answers.

THE *DAUBERT* TRILOGY AND THE FEDERAL RULES OF EVIDENCE

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), the Supreme Court considered the question of the proper interpretation of Federal Rule of Evidence 702 (see p. 12), the principal rule that governs expert testimony. As an initial matter, the Court held that Rule 702 had not incorporated the *Frye* test into the federal practice. The *Daubert* Court held that, unlike under *Frye*, judges could not simply defer to respective fields when assessing the validity of proffered expert testimony. Trial judges are gatekeepers who have the responsibility to assess for themselves the reliability and validity of the methods and principles underlying expert evidence. A judge's responsibility thus changed from merely assessing agreement in a field—the test under *Frye*—to evaluating the research methods and statistics that lay behind expert opinion. Therein lay the revolution of *Daubert*.

The Operating Premises of *Daubert*

Like many Supreme Court decisions, *Daubert* established a foundation upon which a new edifice could be built. *Daubert*, in time, is likely to become more important for the principle it represents than for what it actually says. That principle, in short, is that the law must join the scientific age.

The very first question raised by scholars following the *Daubert* decision was whether the opinion would lead to more expert testimony or less (Chesebro, 1994). This is a debate that has not fully disappeared (Owen, 2002), though most observers find *Daubert* to be fairly restrictive in practice. But if there remained any doubt about the philosophical tenor of *Daubert*, three subsequent Supreme Court decisions dispelled it. The first two join *Daubert* to make up the often-cited "trilogy" of scientific evidence cases. The third is a sometimes-overlooked decision interpreting the scope and import of *Daubert*.

In 1997 the Court ruled in *General Electric Co. v. Joiner* that the Eleventh Circuit Court of Appeals had erred in reversing a trial court's exclusion of expert testimony. Chief Justice Rehnquist wrote for a unanimous Court, holding that appellate courts owe deference to a trial court's admissibility rulings. The Eleventh Circuit had thus erred in second-guessing the trial court's determination that the proffered expert testimony was unreliable. Of great significance, Rehnquist, who had worried over judges playing amateur scientists in *Daubert*, now carefully scrutinized the proffered expertise and, in some detail, pointed out its many deficiencies. Moreover, the Court revised Justice Blackmun's statement that the trial court's obligations were restricted to methods and principles and now made clear that conclusions, too, were part of the court's gatekeeping duties. As Rehnquist put it, "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the ipse dixit of the expert" (p. 146), that is, by the expert's unsupported assertion.

Two years later, in *Kumho Tire Ltd. v. Carmichael* (1999), the Court closed the biggest loophole left from *Daubert*. Justice Stephen Breyer held that a court's gatekeeping responsibilities extended to all expert opinion, not just the scientific variety. According to the Court, therefore, a judge's gatekeeping responsibility includes assessing the validity of experience-based experts as well. The *Kumho Tire* opinion is replete with statements regarding a trial court's abilities to become sophisticated consumers of science and the resources available, such as court-appointed experts, who could assist them in their assigned task.

Finally, in *Weisgram v. Marley* (2000), the Court affirmed an Eighth Circuit opinion that reversed a district court's admission of expert testimony and then held that the expert testimony was not admissible as a matter of law and dismissed the suit. The district court had admitted the testimony of a fire captain, a fire investigator, and a metallurgist in a civil action involving whether a home heater had been defective and had caused the fire. The Court of Appeals reversed. Instead of remanding for a new trial, however, the Eighth Circuit held, as a matter of law, that the expert opinion was inadmissible and then remanded and ordered the district court to enter judgment for the defendant. The Supreme Court affirmed the dismissal of the suit. The Court thus sustained the remarkable power of appellate courts to find expert testimony inadmissible as a matter of law and to direct a verdict accordingly. In agreeing that the plaintiffs should not be given a second chance to find an admissible expert, Justice Ruth Bader Ginsburg stated (pp. 456–457):

> Since *Daubert*, . . . parties relying on expert evidence have had notice of the exacting standards of reliability such evidence must meet. It is implausible to suggest, post-*Daubert*, that parties will initially present less than their best expert evidence in the expectation of a second chance should their first try fail. We therefore find unconvincing [the plaintiff's] fears that allowing courts of appeals to direct the entry of judgment for defendants will punish plaintiffs who could have shored up their cases by other means had they known their expert testimony would be found inadmissible.

Daubert in Practice

The *Daubert* test applies to all types of psychological knowledge, whether based on clinical experience or on extensive empirical testing in the laboratory or in the field. A basic lesson of *Kumho Tire* is that *all* expert evidence—whether it is described as "scientific" or otherwise—must pass a basic threshold test of validity. This is not to say that all expertise is, or is expected to be, comparable. But under *Daubert*, the courts need not make bright-line distinctions among kinds of knowledge. Federal Rule of Evidence 702 speaks in terms of scientific, technical, or specialized knowledge, with all three being subject to the *Daubert* requirement that the judge act as a gatekeeper. But, as the *Kumho Tire* Court recognized, no bright lines divide these three kinds of expert opinions. In fact, all expert opinion is actually based on specialized knowledge (Faigman, 2002). *Daubert* requires judges to assess the worth of such knowledge.

In light of the multitude of contexts in which *Daubert* is employed, from accident reconstruction to nuclear physics, it is not surprising that commentators' views about how the rule applies should vary widely. Nonetheless, there are certain principles that virtually all *Daubert* watchers, whether as supporters or detractors, would agree upon. At its most basic, *Daubert* requires the proponent of expert testimony to bear the burden of proof to show that his or her expert is qualified, that the proffered testimony is relevant to a fact in dispute, and that the basis for the expert's opinion is reliable and valid. In this section we consider these three criteria in regard to psychological expertise.

Qualifications

Rule 702 of the Federal Rules of Evidence speaks of expertise in broad terms. The rule contemplates a wide assortment of potential experts and does not contain degree requirements. An expert must be qualified only by virtue of knowledge, skill, experience, training, or education. The case law, therefore, is replete with statements of deference to proffered expert's qualifications, recognizing the value of experience and practical training.

Cases that exclude experts based on qualifications fall generally into two overlapping camps. In one, the expert lacks the basic credentials or experience to testify on the subject for which he or she is offered. An example might be a nonlicensed therapist seeking to testify regarding the competence of a defendant to stand trial. In the second category, the expert may be well qualified in one area or as a generalist, but has insufficient expertise to testify on the specific subject in dispute. This latter category is by far the more prevalent and also the more controversial.

Some courts adopt a permissive approach to qualifying experts and hold that once the proponent demonstrates the expert's general competency to testify, questions regarding expertise in a specific subdiscipline are a matter of weight and not admissibility. In *Zemaitatis v. Innovasive Devices, Inc.* (2000, p. 633), for example, the court permitted the testimony of a physician who was a jack-of-all-trades. However, a large number of courts agree that a witness may be qualified as an expert on certain matters and not others (*United States v. Roldan-Zapata,* 1990, p. 805). These courts require that the expert display the knowledge or skill necessary to address the specific subarea that is the subject of the litigation (Faigman et al., 2007). Although examples can be found in many areas, courts have particularly frowned upon medical doctors who are offered to testify on every conceivable medical question (e.g., *Alexander v. Smith & Nephew*, 2000).

Psychology presents a similar situation to medical science, in that many practitioners are trained as generalists but there are a number of subdisciplines in which psychologists can specialize. On the whole, however, this issue appears to arise in the courts less frequently in regard to psychologists. One reason may be that in medicine, doctors typically specialize as a matter of residency training, and they rarely jump specialties. An oncologist is likely to have little claim to be an expert on podiatry. If one does make such a claim, courts notice. In psychology, in contrast, specialty areas tend to be less rigidly defined, and generalists often engage in a variety of specialties. Contemporary psychology is primarily defined by either theoretical orientation or developmental stage of the subjects of study—categories that tend to cut across legal issues. Thus, it would not be unusual to see the same behavioral clinician testify in both a competency hearing and a parole determination.

As psychology continues to mature as a discipline, it might very well follow the medical profession in establishing areas of specialization with accompanying certification. This may be especially the case in forensic psychology, which the American Board of Professional Psychology now recognizes as a specialty area of practice and designates educational and training prerequisites. Over time, courts should be expected to become more demanding of the experts who appear before them, insisting on expertise in the specific area of their proposed testimony. In Virginia, for example, an expert may assist a defendant in civil commitment proceedings under the state's Sexually Violent Predators Act on subjects relating to the defendant's mental health (Va. Code § 37.2-907[A], 1999). Under the law, such an expert "shall be a licensed psychiatrist or a licensed clinical psychologist who is skilled in the diagnosis and treatment of mental abnormalities and disorders associated with sex offenders." The Virginia Supreme Court, in fact, recently found that the trial court erred when it admitted the testimony of a psychiatrist who did not treat sex offenders (*Commonwealth v. Miller*, 2007).

The more likely qualifications debate to be heard in courtrooms today involves what credentials should qualify someone to testify as a psychologist. A wide assortment of psychology-related degrees are available, including (but not limited to) the PhD, EdD, PsyD, MS, and MA. Complicating matters further, state certification standards vary widely and often depend on supervised training more than academic education. In general, courts have been permissive in allowing specialists in human behavior to testify. Experience alone often appears to be sufficient to qualify an expert. This has been true, for example, for therapists seeking to testify on the rape trauma

syndrome and the battered woman syndrome (Faigman et al., 2007). It is important to note, however, that qualifying an expert is merely the first step in determining admissibility. Whereas experience alone may be enough to meet the qualifications requirement, it may not be sufficient to validate the substantive aspects of the proposed testimony.

Relevance or Fit

A prerequisite for all admitted evidence is that it must be relevant to some fact in dispute. Although no single piece of evidence is expected to create the chain of inferences necessary to construct a case, every piece of evidence must at least constitute a link in the chain. The same is true for expert testimony. The *Daubert* Court described this requirement as an issue of "fit." The methodological basis underlying an expert's testimony must allow him or her to offer an opinion on a subject in dispute at trial.

The most basic query concerning relevance asks whether the evidence permits an inference to be drawn regarding some fact in dispute. For instance, expert testimony that a defendant lacks volitional control over his or her behavior is relevant in a state employing the American Law Institute (ALI) test of insanity but is irrelevant in states using the M'Naughten test, under which the defendant must prove that he or she cannot distinguish right from wrong. Relevance ordinarily is a basic logical proposition that asks whether knowing X makes Y more or less likely, where Y is a fact that is material to the law that applies to the case. Under the ALI test for insanity, lack of volitional control is an element of the defense, but under M'Naughten, it is not.

In the realm of expert evidence, the fit inquiry raises methodological issues regarding whether the methods used in the research underlying the disputed expert evidence are sufficient to support the opinion offered in a particular case. This is primarily an issue of external validity. A classic instance of this point concerns whether research on the toxicity of certain substances generalizes from animal research to humans. Psychological research is replete with such issues. Does jury research on undergraduates generalize to real jurors? Would a finding of discrimination in one state's capital sentencing system permit inferences about another state's process? Does research on children's competency in one legal context generalize to other legal contexts? Because so much psychological research uses subjects or situations that depart widely from the legal realities, this aspect of relevancy can present substantial obstacles to the admission of psychological opinion.

Reliability and Validity

The core holding of *Daubert* was the Court's interpretation of Federal Rule of Evidence 702 that the proponent of expert testimony has the burden to demonstrate by a preponderance of the evidence that the basis for the proffered opinion is valid. Although the *Daubert* Court stated that it was merely interpreting the plain meaning of the Federal Rules, the rules themselves were amended in 2000 to reflect the meaning the Court had found in them 7 years before. The new Rule 702 essentially codified *Daubert*:

> 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.

The *Daubert* Court suggested four factors that courts might use to assess the scientific validity underlying expert testimony. These were (1) testing, (2) error rate, (3) peer review and publication, and (4) general acceptance. In addition, courts and commentators have usefully suggested additional factors that might be employed to complete the assigned task (Faigman et al. 2007). Most importantly, courts have increasingly recognized that the myriad kinds of expertise they confront require a flexible and nuanced approach. Hence, peer review and publication may be an appropriate criterion for both histologists and historians, but error rate is not. *Daubert* mandates that trial courts determine whether the basis for proffered expert opinion is reliable and valid. The process trial courts use to reach this determination can be adjusted for each category of expertise and is largely within each court's discretion. At the same time, it is clear that some process and some factors must be employed to make this judgment. Justice Antonin Scalia emphasized this point in his *Kumho Tire* (1999) concurrence. He stated, "Although, as the Court makes clear today, the *Daubert* factors are not holy writ, in a particular case the failure to apply one or another of them may be unreasonable and hence an abuse of discretion" (p. 159).

Since psychological testimony is ostensibly based on scientific knowledge, the *Daubert* factors are a reasonable starting point in most cases. We therefore use the *Daubert* factors below to frame our discussion of psychological expertise.

TESTING

Although the *Daubert* Court listed testability as one of its four basic factors, it is almost certainly a prerequisite to all that follows. In psychology, as in all science, without testing, error rates could not be ascertained, peer-reviewed journals ordinarily would not publish the work, and general acceptance would not be forthcoming. It is important to remember that *Daubert* requires not merely that the basis for the opinion be testable, but that it has been adequately tested (Black, Ayala, & Saffran, 1994). Ultimately, the challenge of *Daubert* lies in measuring this adequacy.

Properly understood, the *Daubert* factors represent a set of criteria by which fairly traditional scientific evidence might be tested. Although the actual inquiry might be significantly more robust, scientists themselves would consider whether and how a hypothesis had been tested, the degree and type of error associated with the science or technology, whether the report of the research was published in a high-quality, peer-reviewed journal, and what scientists in the respective or associated fields thought about the work. When an expert proposes to testify regarding knowledge that is amenable to evaluation by those criteria, courts should use those criteria to assess the success of that knowledge.

ERROR RATE

The *Daubert* Court's use of error rates is at once completely understandable and quite perplexing. Error, of course, is a core component of all science, and it comes in an assortment of varieties. Moreover, from the judicial perspective, the amount of potential error expert evidence poses ought to affect judgments about its admissibility. The cost of making a mistake, whether of the false-positive or false-negative variety, is an integral component of the policy implications of any admissibility determination. Yet Justice Blackmun's mention of error rate was perfunctory and not well designed to give most experts substantial guidance. He merely stated that, "in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error" (*Daubert*, p. 594). He thus appeared to limit his concern with error to techniques such as DNA profiling or polygraph tests, which might have an identifiable error rate with repeated applications. Lower courts since *Daubert* have done little to remedy the deficiencies in Blackmun's analysis of error rates. For instance, courts rarely distinguish between Type I and Type II error when they discuss this factor. Error is a multidimensional construct with widely varying implications for the law. It

is worthy of much more attention than it has so far received from the courts and, indeed, is a subject deserving of sustained scholarly attention.

PEER REVIEW AND PUBLICATION

In *Daubert*, the Court cited peer review and publication as a helpful proxy of good-quality research, but emphasized that it should not be considered the sine qua non of validity. Although publication in peer-reviewed journals is the hallmark of the successful completion of research, the process does not guarantee quality data. Much that is published in the very best journals can be criticized, is impervious to replication, or turns out to be invalid (Jasanoff, 1996). Also, some work that appears outside the strict scientific peer-review process is well done and well worth relying on. Nonetheless, the ordinary culmination of most scientific research is publication in a peer-reviewed journal, and, as implied in *Daubert*, expert opinion that comes without this imprimatur is suspect.

Also, a somewhat less noticed aspect of the peer-review and publication factor is the more general dynamic present in healthy scientific communities. Psychology is somewhat less successful at this more general checking function than many sciences. The single most effective checking tool in science is replication. If one laboratory announces the discovery of cold fusion, for instance, others inevitably follow fast on its heels to assess the validity of the claim. This is true even with many complex applied phenomena, such as the epidemiology of the effects of some alleged toxic substance or product. One or two studies of a phenomenon from one or two laboratories are ordinarily not enough to recommend policy. In psychology, in contrast, it is not unusual for one or two studies to serve as the substantive foundation for policy recommendations.

GENERAL ACCEPTANCE

Like peer review and publication, the use of widespread or general acceptance as a criterion of validity depends on the quality of the field from which the findings come. Unlike the testing and error-rate factors, general acceptance is merely a proxy for validity and is only as good as those doing the accepting or rejecting. The biggest danger associated with this factor from the law's standpoint is that consensus might replace critical assessment. This has largely occurred, for example, in the forensic specialties, such as bite-mark and handwriting identification analysis, where the main community involved is law enforcement and dissent is strongly frowned upon. Although

psychology, as a field, demonstrates much greater critical judgment than many forensic areas, it nevertheless shares some of the difficulties that are endemic in the forensic specialties. Especially deleterious to self-criticism is that some psychologists measure the success of their work by whether courts accept or reject it.

The *Daubert* and *Frye* tests both create fairly substantial hurdles for expert evidence to overcome. Many states, however, create an exception to their applicable expert testimony rule for expert opinion that is not scientifically derived. In these jurisdictions, experience-based expert opinion must pass only a stunted relevancy inquiry.

THE RELEVANCY TEST

The relevancy test in many respects constitutes the foundational requirement for all evidence under virtually every evidence code. Under the test, the proponent is first obligated to demonstrate that the evidence has probative value regarding an issue in dispute under the substantive law that applies in the case. Typically, this simply means that the evidence makes a material fact in the case more likely or less likely to be true. It is a modest requirement. In addition, the use of the relevancy test in the context of expert testimony ordinarily incorporates a competency of the witness requirement. In the case of experts, this involves the question whether the expert is qualified. Where the relevancy test is employed, however, qualifications are usually interpreted broadly, and doubts about an expert's credentials are typically left to the jurors to resolve in evaluating the weight of the evidence.

Although a few states, such as Wisconsin, use a relevancy test exclusively for expert testimony, it is widely employed as an exception to the prevailing rule for expert testimony. California and Arizona, for example, both exempt nonscientific, experience-based, expert opinion from their *Frye*-based tests. This practice disproportionately affects medical and psychological expert testimony. For instance, such prominent subjects as repressed memories, predictions of violence, posttraumatic stress disorder, and child abuse accommodation syndrome are not subject to a *Frye* analysis (Faigman et al., 2007). This section considers whether the basis for this exception is well founded.

California divides the world of expert evidence into two basic categories. On the one hand, experts sometimes offer opinions that are based on scientific or technical processes or tests, whereas, on the other hand, experts sometimes offer opinions based on professional experience. The former tes-

timony is thought to be particularly worrisome because of the aura of certainty that might surround opinion based on the seeming wizardry of science. Using *Frye*, California erects evidentiary barriers to this sort of expert opinion, fearing that it might overwhelm the tender sensitivities of fact-finders. In contrast, expert opinion that is based on experience and claims no pretensions to scientific exactitude does not receive this close evidentiary review. In effect, California exempts experiential expert opinion—or non-scientifically derived opinions—from the rigors of evidentiary review. This practice, one employed explicitly or implicitly in many jurisdictions, is often referred to as the "opinion rule." The opinion rule is supported neither by jurisprudential nor empirical considerations.

Consider the case of *People v. Miller* (2005), in which a California appellate court applied the opinion rule and held that psychiatric and psychological testimony predicting future violence under the state's Sexually Violent Predator Act is not scientific evidence subject to California's version of the *Frye* test, termed *Kelly–Frye* (*People v. Kelly*, 1976). The defendant had complained that the state's expert had relied on statistical tests that should have been demonstrated to be generally accepted by the relevant scientific community. The *Miller* court, however, rejected this argument, pointing out that the expert's opinion was clinically based and the statistical tests were used, at most, to support an experience-based clinical judgment. The court explained "that since the expert's opinion testimony was not based solely on the actuarial evidence, but rather on a variety of factors and the expert indicated such evidence was not infallible, a *Kelly* hearing was not required" (p. 1). The California court, therefore, sought to draw a bright-line distinction between actuarial techniques, which are derived from extensive research programs, and clinical judgments, which are primarily experience-based determinations. In practice, risk assessments exist on a continuum of structure, with completely unstructured (corresponding to "clinical") assessment occupying one end of the continuum and completely structured assessment (corresponding to "actuarial") occupying the other, but with additional forms of more-than-unstructured-but-less-than-fully-structured risk assessments lying between these poles (Monahan, 2008).

Under the California approach to expert testimony, therefore, clinical psychological opinion that is not primarily based on statistical techniques is admitted with no *Kelly–Frye* threshold check, but psychological opinion that *is* premised on such technology confronts the formidable barriers of that test. In effect, the California rule means that expert opinion with little or no scientific basis is readily admitted, but opinion that is based on scientific test must survive the *Kelly–Frye* gauntlet. This approach is particularly

perverse in the context of violence risk assessment, since research has consistently demonstrated over the last five decades that well-validated empirically structured risk assessments are much more accurate than unstructured clinical judgment (Meehl, 1954; Monahan, 2007b). Moreover, research in the last decade is leading to the development of ever more powerful actuarial technology, so the gap between clinical judgment and actuarial methods is likely to further widen (Monahan, 2007a). Yet the California approach effectively gives a free pass to experience-based clinical judgment and erects substantial barriers to the introduction of science-based actuarial techniques.

California and other jurisdictions advance two basic arguments in support of the rule that relaxes evidentiary standards for experience-based expert opinion. The first is not always spelled out, but such a regime must presume that a dividing line can be identified between scientific and nonscientific evidence. The second argument advanced to support the California rule is the belief that jurors are more likely to be impressed by the aura of infallibility that surrounds scientific opinion, but can critically assess nonscientific opinion readily enough.

DISTINGUISHING SCIENCE FROM NONSCIENCE

Although there is certainly a distinction to be drawn between science and nonscience, logic does not recommend its use to support a rule that would allow nonscientific opinion easy admission and enact substantial hurdles to the admission of science-based evidence. That something is not science is hardly reason to admit it for the jury's consideration. The California courts have not explored this issue in any depth, but have merely assumed that a line dividing science and nonscience is readily identifiable. Other courts applying the opinion rule have sought to define the parameters of the boundary between science and nonscience.

The most ambitious attempt to set forth the premises underlying the opinion rule was advanced by the Arizona Supreme Court in *Logerquist v. McVey* (2000). The court held that the *Frye* test—the state's ordinary rule of admissibility for scientific evidence—did not apply to nonscientific expert opinion. The court explained, "*Frye* is applicable when an expert witness reaches a conclusion by deduction from the application of novel scientific principles, formulae, or procedures developed by others" (p. 133). However, the court argued, *Frye* "is inapplicable when a witness reaches a conclusion by inductive reasoning based on his or her own experience, observation, or research" (p. 133). Hence, for the *Logerquist* court, opinion based on induc-

tive reasoning is not subjected to any substantial threshold test, but opinion that stems from deductive reasoning receives close scrutiny.

The error the *Logerquist* court makes is believing that scientific knowledge is discrete or categorical. It is not. The scientific method is a process. Indeed, it is a process that invariably begins with experience. Consider the example of repressed memories, the subject of the *Logerquist* decision. A doctor might have examined one or more people who claim to have begun remembering things of a disturbing nature and of which they had been previously unaware. A scientist's approach to this observation would be at least twofold. On the one hand, employing inductive reasoning, a scientist might look for further instances of this phenomenon. The initial observation could have been an anomaly or not accurate for a variety of reasons. In seeking confirmation, the wise researcher would want to ensure that the underlying traumas occurred, that the amnesia was not explainable by other factors—such as biological amnesia—and that the reports of repression were not spurious products of expectations of either the subjects or researchers. At the same time, using deductive reasoning, the scientist would seek to integrate his or her observations into what is generally known about human memory and the brain. Does this observation conform to theory? Perhaps theory suggests another explanation for the recalled memories other than repression. Or possibly—and of potentially great excitement for the scientist—the theory itself needs to be adjusted to account for these newly discovered facts.

Given the structure of science, the distinction the *Logerquist* court makes between deductive and inductive reasoning, for purposes of admissibility standards, is chimerical. What the court called inductive reasoning is simply hypothesis formation through anecdotal experience. The subject of repressed memories well illustrates the point. In *Logerquist*, the plaintiff's expert, Dr. van der Kolk, said that he had treated "hundreds of survivors of childhood sexual abuse" and that he was prepared to testify that some of these "victims do have delayed memories, that their memories are as reasonably accurate as normal memories, if not better" (p. 117). How would he know something like this? Even more arrogant is the doctor's claim that "delayed memories" are as "accurate as normal memories, if not better" (p. 117). Better? He thus claims experience with repressed memory and normal memory formation, comparisons of the two, as well as the investigatory capacity to check claims of abuse several decades old. That is some very impressive experience.

The basic error of *Logerquist* lies in the court's failure to critically assess the proffered evidence in a scientifically sophisticated way. It is not simply that repression is not yet science, it is that repression has repeatedly failed scientific attempts to validate it (Faigman et al., 2007).

JUROR CREDULITY

Even if a realm of nonscientific expert opinion could be identified reliably, is it reasonable to assume that no substantial threshold test is necessary because jurors can readily assess its validity and weight? According to the California Supreme Court, the primary concern with scientific opinion is the possibly overwhelming influence complex scientific evidence has on jurors (*People v. Kelly,* 1976). The court observed that "lay jurors tend to give considerable weight to 'scientific' evidence when presented by 'experts' with impressive credentials" (p. 1244). The *Frye* threshold, therefore, establishes a barrier in order to protect against juror credulity. Nonscientific expert opinion, according to the California high court, does not present a similar danger: "When a witness gives his personal opinion on the stand—even if he qualifies as an expert—the jurors may temper their acceptance of his testimony with a healthy skepticism born of their knowledge that all human beings are fallible" (*People v. McDonald*, 1984, p. 724).

A basic premise implicit in the California approach is the apparent belief that jurors are better able to parse personal opinion than they are able to critique scientific opinion. The court in *People v. Venegas* (1998) summarized this argument:

> The *Kelly* test is intended to forestall the jury's uncritical acceptance of scientific evidence or technology that is so foreign to everyday experience as to be unusually difficult for laypersons to evaluate. In most other instances, the jurors are permitted to rely on their own common sense and good judgment in evaluating the weight of the evidence presented to them. (p. 546)

Although it might indeed challenge jurors to critically assess scientific opinion, it is not clear that they are likely to do considerably better with nonscientific opinion. How, in fact, should a juror evaluate experience-based expert opinion? The California Supreme Court advances common sense and good judgment as if these characteristics were self-defining and obviously applicable to the task at hand.

Consider again the issue of repressed memories, a subject generally thought to be exempt under the opinion rule (*Wilson v. Phillips*, 1999). An expert might testify that the plaintiff repressed her memory from a young age until adulthood, that such a phenomenon has occurred in other cases, and that the memories recalled are reliable. If you are a commonsensical juror with good judgment, how do you assess this claim? There are an assortment of possibilities, such as demeanor, the prospect of bias due to

expert witness fees, credentials, and admissions or inconsistent statements elicited on cross-examination. Of course, all of these commonsense good judgments are available to critically assess scientific opinion as well. Other than these standard indicia of reliability, all with questionable value, there is nothing more. Nonscientific expert opinion is, in fact, little more than *ipse dixit*. Repressed memories are so because an expert with 20 years of experience with the subject says that they are so. By this measure, bloodletting and alchemy were valid too.

Cross-examination is unlikely to be particularly effective with well-credentialed experts for another reason. Most of these witnesses believe the content of their testimony. A lying witness may have sweaty palms, avoid eye contact, and stutter. A lying witness might also be caught in a contradiction or be unable to explain prior inconsistent statements. But many experience-based experts are fully convinced of the validity of their judgment. Experience, after all, has failed to falsify their beliefs. The weakness of their testimony does not lie with the witness, it lies with the content of their opinion and the methods they used to form it.

California's practice of employing a restrictive test for science-based expert opinion and a permissive test for experience-based expert opinion is arguably the opposite of what good scientific common sense would recommend. At the very least, however, there is no compelling basis for California's approach of using two different tests for expert opinion evidence. Expert opinion, whatever the ostensible foundation it rests upon, presents similar dangers to jurors. Moreover, the factors that are likely to be relevant to determining the weight of proffered expert opinion are generally the same, whether the evidence is scientifically or experientially based.

Perhaps the strongest argument against the exception for experience-based opinions is practical. In *Miller* itself, the expert purported to rely on a combination of actuarial and clinical indicia to reach an opinion. It was the fact that the expert did not rely "solely on the actuarial evidence" that was determinative for the court. In the context of risk assessment, this integration of structured test and clinical experience is generally standard operating procedure. Indeed, it would be difficult to identify any area of applied science in which experience did not buttress an expert's opinion that was ostensibly based on a structured or mechanical test. No expert comes into court and simply recites the outcome of some test, whether it is a radiologist "interpreting" an x-ray or a polygrapher reporting the "results" of a polygraph. Moreover, even if a scientific test provided "the" answer, inevitably some experience-based judgments would have been built into the test. Every test has a decision threshold that is associated with human judgment

somewhere along the way (Swets, Dawes, & Monahan, 2000). California's attempt to draw distinctions between "science" and "nonscience" is doomed to fail. The rational approach in the risk assessment area, therefore, would be to hold *Kelly–Frye* hearings for (1) purely actuarial prediction, (2) clinically reviewed actuarial prediction, and (3) purely clinical prediction.

FUTURE PROSPECTS

Greater Legal Sophistication through Court-Appointed Experts

As we noted at the start of this chapter, the revolution sparked by *Daubert* primarily involved a changed perspective. Under *Frye*'s general acceptance test, admissibility determinations consisted of little more than counting noses of professionals in a particular field. *Daubert* demands that judges evaluate the research methods supporting expert evidence and the principles used to extrapolate from that research to the task at hand (Risinger, 2007). This is a daunting task, especially for judges who, on average, have little formal training in statistics and research methods. How, it may reasonably be asked, are judges to become sophisticated consumers of psychological and other expertise, as *Daubert* requires?

One proposed solution to this question is greater use of court-appointed, rather than party-retained, experts. Scholars have actively advocated this device (Berger, 1994; Gross, 1991). Although research indicates that courts are disinclined to appoint experts for themselves, there are signs that this reluctance is slowly changing (Faigman et al., 2007). Cecil and Willging (1994) found that judges relied little on this appointment power and, moreover, most (62%) of those surveyed thought that the "appointment of an expert [is] an extraordinary action" (pp. 1015–1018). The principal reason for this view was judges' belief that court-appointed experts undermine the adversarial process. Yet, as experience with the challenges of implementing *Daubert* has been gained, courts appear increasingly sympathetic to the possibility of appointing experts (*Hall v. Baxter Healthcare Corp.*, 1996). Indeed, Justice Breyer, concurring in *General Electric Co. v. Joiner* (1997), quoted approvingly the *New England Journal of Medicine*'s amicus brief calling for greater use of court-appointed experts:

> [A] judge could better fulfill this gatekeeper function if he or she had help from scientists. Judges should be strongly encouraged to make greater use of their inherent authority … to appoint experts … . Reputable experts could be recommended to courts by established scientific organizations, such as the

> National Academy of Sciences or the American Association for the Advancement of Science. (p. 149)

In federal cases, court experts assume primarily one of two forms. The first, properly termed court-appointed experts, are selected mainly to assist jurors determine what weight to give the expert evidence that is presented by the parties. These experts give testimony in court, are subject to the ordinary procedures of discovery, and may be cross-examined by one or both parties. The second type of court expert, increasingly prevalent in the case law, is the technical adviser. These experts' primary function is to assist judges rather than juries. They typically are not subject to adversarial processes such as discovery or cross-examination. In effect, technical advisers sit at judges' sides, like law clerks, assisting them to understand the expert evidence that the parties seek to have admitted at trial. In the future, greater use of psychologists as court-appointed experts or as technical advisors could go far in educating juries and judges to be sophisticated consumers of psychological evidence.

Greater Psychological Sophistication through Evidence-Based Practice

We write at a time when a debate rages in clinical psychology around what is being termed "empirically validated treatment" or, more broadly, "evidence-based practice." On one side of this debate are Scott Lilienfeld (2002), David Barlow and colleagues (Barlow, Levitt, & Bufka, 1999), and Richard McFall (1991), clinicians themselves, who take the view that clinical psychologists should restrict their professional activities to those that have ample support in the scientific literature. As Lilienfeld has stated:

> Once we abdicate our responsibility to uphold scientific standards in administering treatments, our scientific credibility and influence are badly damaged. . . . Our students will most likely follow in our footsteps and continue to turn a blind eye to the widening gap between scientist and practitioner, and between research evidence and clinical work. (p. 9)

On the other side of this debate are Ronald Fox (2000) and Ronald Levant (2004), both of whom are past presidents of the American Psychological Association. Fox (pp. 1–2) has written, "Psychologists do not have to apologize for their treatments. Nor is there any actual need to prove their effectiveness." Levant has taken the position that in clinical psychology

"many treatments have not been empirically studied, and that there is a big difference between a treatment that has not been tested empirically and one that has not been supported by the empirical evidence" (p. 221). An emphasis on evidence-based practice, Levant believes, has already hurt practitioners by restricting the interventions that are reimbursed by insurance companies, and this emphasis "could create additional hazards for practitioners in the courtroom if empirically-validated treatments are held up as the standard of care in our field" (p. 221). Rather than reliance on evidence-based practice, Levant has called for "a broad perspective that allows the integration of the research (including that on empirically validated treatments . . .) with clinical expertise and, finally, brings the topic of patient values into the equation. Such a model, which values all three components equally, will better advance knowledge related to best treatment and provide better accountability" (p. 223).

We would agree with the views of another clinician, however. As Larry Beutler (2004) has stated, "To elevate even the best clinical judgment to the status of knowledge, rather than hunch . . . is the point at which we will both jeopardize our status among learned professions and harm our patients. . . . Contrary to Levant, research, experience, and patient values are not equivalently valid. Scientific research is more likely to produce valid conclusions than sincere clinical opinion based on unsystematic experience" (p. 228).

Rarely has the law come down so forcefully on what might be seen as an intraprofessional dispute (Appelbaum, 2002). *Daubert* unequivocally endorses "empirically validated treatments" and "evidence-based practices." Under *Daubert*, there is, in fact, *no difference* "between a treatment that has not been tested empirically, and one that has not been supported by the empirical evidence" (Levant, 2004, p. 221). Neither is admitted. Whatever clinical value unvalidated psychological assessment or treatment techniques may or may not have in the consultation room, *Daubert* makes plain that testimony employing such techniques has no place on the witness stand.

REFERENCES

Alexander v. Smith & Nephew, 90 F.Supp.2d 1310 (N.D. Okla. 2000).

Appelbaum, P. S. (2002). Policing expert testimony: The role of professional organizations. *Psychiatric Services, 53*(4), 389–399.

Barlow, D., Levitt, J., & Bufka, L. (1999). The dissemination of empirically supported treatments: A view to the future. *Behavioral Research and Therapy, 37,* 147–162.

Berger, M. (1994). Novel forensic evidence: The need for court-appointed experts after *Daubert*. *Shepard's Expert and Scientific Evidence, 1*, 487–512.

Beutler, L. E. (2004). The empirically supported treatments movement: A scientist-practitioner's response. *Clinical Psychology: Science and Practice, 11*(3), 225–229.

Black, B., Ayala, J., & Saffran, C. (1994). Science and the law in the wake of *Daubert*: A new search for scientific knowledge. *Texas Law Review, 72,* 715–802.

Cecil, J., & Willging, T. (1994). Accepting *Daubert*'s invitation: Defining a role for court-appointed experts in assessing scientific validity. *Emory Law Journal, 43*, 995–1070.

Chesebro, K. J. (1994). Taking *Daubert*'s "focus" seriously: The methodology/conclusion distinction. *Cardozo Law Review, 15,* 1745–1753.

Commonwealth v. Miller, 643 S.E.2d 208 (Va. 2007).

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Faigman, D. L. (1989). To have and have not: Assessing the value of social science to the law as science and policy. *Emory Law Journal, 38,* 1005–1095.

Faigman, D. L. (2002). Is science different for lawyers? *Science, 197*, 339–340.

Faigman, D. L., Kaye, D. H., Saks, M. J., Sanders, J., & Cheng, E. K. (2007). *Modern scientific evidence: The law and science of expert testimony* (2d ed.). Minneapolis, MN: West/Thompson.

Faigman, D. L., Porter, E., & Saks, M. J. (1994). Check your crystal ball at the courthouse door, please: Exploring the past, understanding the present, and worrying about the future of scientific evidence. *Cardozo Law Review, 15,* 1799–1835.

Fox, R. E. (2000). The dark side of evidence-based treatment. *Practitioner Focus*. Retrieved April 27, 2008, from *www.apa.org/practice/pf/jan00/cappchair.html*.

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

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior, 25,* 433–458.

General Electric Co. v. Joiner, 522 U.S. 136 (1997).

Gross, S. (1991). Expert evidence. *Wisconsin Law Review*, pp. 1113–1232.

Hall v. Baxter Healthcare Corp., 947 F.Supp. 1387 (D.Or. 1996).

Jasanoff, S. (1996). *Science at the bar: Law, science, and technology in America*. Cambridge, MA: Harvard University Press.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56,* 405–416.

Kumho Tire v. Carmichael, 526 U.S. 137 (1999).

Levant, R. (2004). The empirically validated treatments movement: A practitioner/educator perspective. *Clinical Psychology: Science and Practice, 11*, 219–224.

Lilienfeld, S. O. (2002). The scientific review of mental health practice: Our raison d'etre. *Scientific Review of Mental Health Practice, 1*, 5–10.

Logerquist v. McVey, 1 P.3d 113 (2000).
McFall, R. M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist, 44*, 75–88.
Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
Monahan, J. (2007a). The scientific status of research on clinical and actuarial predictions of violence. In D. L. Faigman, D. H. Kaye, M. J. Saks, J. Sanders, & E. K. Cheng (Eds.), *Modern scientific evidence: The law and science of expert testimony* (2nd ed., pp. 124–149). Minneapolis, MN: West/Thompson.
Monahan, J. (2007b). *Tarasoff* at thirty: How developments in science and policy shape the common law. *University of Cincinnati Law Review, 75*, 497–521.
Monahan, J. (2008). Structured violence risk assessment. In R. Simon & K. Tardiff (Eds.), *American Psychiatric Publishing textbook on violence assessment and management* (pp. 17–33). Washington, DC: American Psychiatric Publishing.
National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: National Academies Press.
Owen, D. G. (2002). A decade of *Daubert*. *Denver University Law Review, 80*, 345–372.
People v. Kelly, 549 P.2d 1240 (1976).
People v. McDonald, 690 P.2d 709 (1984).
People v. Miller, 2005 WL 768749 (Cal.App.4 Dist. 2005).
People v. Smith, 784 N.Y.S.2d 923 (Table), 2004 WL 690321 (N.Y. Supp. 2004).
People v. Venegas, 954 P.2d 525 (1998).
Risinger, D. M. (2007). Preliminary thoughts on a functional taxonomy of expertise for the post-*Kumho* world. In D. L. Faigman, D. H. Kaye, M. J. Saks, J. Sanders, & E. K. Cheng (Eds.), *Modern scientific evidence: The law and science of expert testimony* (2nd ed., pp. 119–151). Minneapolis, MN: West/Thompson.
Sexually Violent Predator Act, VA. Code § 37.2-907 A (1999).
Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decision. *Psychological Science in the Public Interest, 1*, 1–26.
United States v. Roldan-Zapata, 916 F.2d 795 (2d Cir. 1990).
Weisgram v. Marley Co., 528 U.S. 440 (2000).
Wilson v. Phillips, 86 Cal.Rptr.2d 204 (4th Dist. 1999).
Zemaitatis v. Innovasive Devices, Inc., 90 F.Supp.2d 631 (E.D.Pa. 2000).