PSYCHOLOGICAL EVIDENCE AT THE DAWN OF THE LAW’S SCIENTIFIC AGE

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

John Monahan
School of Law, University of Virginia, Charlottesville, Virginia 22903; email: jmonahan@virginia.edu

Key Words evidence, testimony, law, expert, Daubert

Abstract The Supreme Court’s 1993 decision in Daubert v. Merrell Dow Pharmaceuticals, Inc., holding that the admissibility of scientific evidence depends on its scientific merit, has made American law receptive to valid science to an unprecedented degree. We review the implications for psychology of the law’s taking science seriously. We consider the law before Daubert, and the ways that Daubert as well as modifications to the Federal Rules of Evidence have affected the admissibility of expert testimony. We describe the ramifications of these changes for psychology used as authority to create a general legal rule, as evidence to determine a specific fact, and as framework to provide context. Finally, future prospects for improving psychological testimony are offered: Court-appointed experts will increase the psychological sophistication of judges and juries, and evidence-based practices on the part of psychologists will increase the sophistication of the expert testimony that they proffer in court.

CONTENTS

INTRODUCTION ....................................................... 632
THE TIME BEFORE DAUBERT ........................................ 633
THE DAUBERT TRILOGY AND THE FEDERAL RULES OF
EVIDENCE .......................................................... 635
The Operating Premises of Daubert ................................. 635
Daubert in Practice .................................................. 638
THE STRUCTURE OF PSYCHOLOGICAL SCIENCE
AND THE RULES OF ADMISSIBILITY .......................... 647
Social Authority ..................................................... 648
Social Facts .......................................................... 650
Social Frameworks ................................................... 652
FUTURE PROSPECTS ................................................ 654
Greater Legal Sophistication Through Court-Appointed Experts ........ 654
Greater Psychological Sophistication Through Evidence-Based Practice .... 655

0066-4308/05/0203-0631$14.00
INTRODUCTION

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), the U. S. Supreme Court held that the admissibility of scientific evidence depends on its scientific validity. The *Daubert* holding was deepened and broadened by subsequent developments, including other Supreme Court decisions and extensive amendments to the Federal Rules of Evidence governing expert testimony. Additionally, the federal courts decide hundreds of cases per year involving *Daubert*-related issues, and the law reviews have produced literally thousands of articles on the subject. The states too are actively involved in this area, with a majority now following a *Daubert* approach and many more deeply influenced by that decision. Psychology is fully caught in the *Daubert* maelstrom. In fact, psychological science offers a particularly revealing window on the entire subject of expert evidence. Because of the many different types of expertise captured within the umbrella of psychology, psychological evidence presents the full range of issues that have bedeviled the courts since *Daubert* decided to take science seriously.

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 Ltd. v. Carmichael* (1999), one of the cases that followed *Daubert*, 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.” However, 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). The bases for expert opinions range from years of experience to rich literatures of empirical research. As a practical matter, science is a method of inquiry, not a category of knowledge. The challenge lies in determining what methods are adequate to support a particular expert opinion.

The approach introduced in *Daubert* changes the focus of the judicial inquiry from the ostensible category the opinion falls into (scientific/nonscientific) to the methods that were employed to evaluate the empirical basis for the opinion. In terms of applied science, no one method is fully privileged, though certain methods may be disfavored because they provide little dependable data. The ultimate legal issue presented is whether the methods are adequate and sufficient to support the expert opinion proffered in court. To this end, the Federal Rules of Evidence and a majority of state evidence codes now require judges, as gatekeepers, to determine whether the basis for expert testimony is more likely than not reliable and valid.

*Daubert* has largely changed the way the field of psychology and law will henceforth conceive of itself. *Daubert’s* emphasis on the scientific method is likely to create a situation of the haves and the have-nots in psychology and the law. Many
areas of psychology—too many to give them their due in this review—readily pass muster under *Daubert*. The most prominent example of this *Daubert* success, though hardly without exception, is the large percentage of the voluminous literature on the reliability of eyewitness testimony (Wells & Olson 2003). But many areas of psychology that are relevant to legal decision making lack robust research traditions and are likely to suffer under the *Daubert* lens. The years ahead will be difficult ones for experts whose opinions rest on shaky empirical foundations. Indeed, *Daubert* is likely to throw fuel on the already raging debate among psychologists regarding what methods are adequate to study human behavior. With the courts now participating, this debate assumes even greater practical significance. In time, as lawyers become more sophisticated in their understanding of science, many areas of psychology, previously allowed easy entry to the courtroom by virtue of general consensus in the profession, will now be stopped and turned away.

The field of psychology and law is inextricably bound to developments in the area of expert evidence (Borgida & Hunt 2003). This review endeavors to describe and analyze these developments. Because the extensive empirical literature on jurors’ understanding of expert evidence has been reviewed recently and comprehensively by Vidmar & Diamond (2001; see also Diamond et al. 2003, Ivkovic & Hans 2003), we do not address this topic. Rather, we first explore the time before *Daubert*, when the “general acceptance” test of *Frye v. United States* (1923) 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*. We then turn more fully to the *Daubert* case and subsequent developments in case law and in amendments to the Federal Rules of Evidence. With these legal developments as background, we examine the basic lessons *Daubert* provides for psychology. At least in theory, if not always in practice, strategies for dealing with one type of expertise ought to apply to analogous kinds of expertise. Lessons for psychology can thus be drawn from the myriad other forms of expertise vying for court’s attention, including chemistry, biology, sociology, and medicine. Finally, we consider the future prospects for psychology and psychologists under *Daubert*. We address procedural matters, such as court-appointed experts and technical advisers, as well as the ways that *Daubert* bears on substantive debates within the field of psychology.

THE TIME BEFORE *DAUBERT*

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 alternative to that test that is most often posited, the general acceptance standard of *Frye v. United States*
(1923) 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 et al. 1994).

Courts’ and scholars’ views of Frye were 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 science. 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. Frye seemed to ensure an orderly and accurate, albeit gradual, development of the jurisprudence of scientific evidence.

Frye also had more than its share of detractors. Some critics charged that the test was too conservative, leading courts to exclude great advances until the respective fields embraced the new work, a process that could take years in some cases. In many applications, therefore, Frye was a very conservative influence. At the same time, Frye was also 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 the use of poor research methods. 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. However, the relevance of the science may be very different for the law than it is for the researchers and practitioners. 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. However, the acceptance of the polygraph as a tool of interrogation says little about its scientific
validity. A National Research Council report, for instance, concluded that poly-
graphs had doubtful validity, especially for national security screening purposes,
but did not reach any conclusions about their utility as props for interrogations
or their deterrent value (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. Judges must ask the right questions under Frye in order to get
useful answers.

In 1993, the Supreme Court considered the question of whether the Federal
Rules of Evidence had incorporated the Frye test when they were enacted in
the 1970s. The Court held that they had not and went on to define the test that
would rock the legal world. In Daubert v. Merrell Dow Pharmaceuticals, Inc.
(1993), the Court held that judges could not simply defer to respective fields when
assessing the validity of proffered expert testimony. Trial judges are gatekeepers
and they have the responsibility to assess for themselves the reliability and validity
of the methods and principles underlying expert evidence (Note 2003). 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 DAUBERT TRILOGY AND THE FEDERAL RULES
OF EVIDENCE

The Operating Premises of Daubert

Among both lawyers and scientists, Daubert may be the best-known Supreme
Court decision outside of constitutional law. It is also one of the least understood.
Blame for the confusion surrounding the decision lies in part in the opinion itself.
Justice Harry Blackmun, Daubert’s author, has generally been considered one of
the more scientifically sophisticated justices and, at first glance, Daubert would
seem to support this view of him. Yet, his opinion contains a hodgepodge of con-
flicting statements, enough so that almost everyone can find something to love and
something to hate in the decision. In addition, the philosophy of science he em-
ployed was deemed reasonable by some, anachronistic by others, and incomplete
by almost everyone (Haack 2003, Sanders et al. 2002). Yet, like many Supreme
Court decisions, it established a foundation upon which a new edifice could be
built. Daubert, in time, is likely to become more important for the principle it
stands for 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 debate has not fully disappeared (Owen 2002), though most observers now
believe that the answer is generally less. Blackmun’s opinion at first extolled the
liberal admissibility standards of the Federal Rules of Evidence. It then adopted the most conservative standard the Rules offered for assessing expert opinion. Blackmun also described the trial court’s role as gatekeeper and explicitly recognized that the rule would lead to the exclusion of certain kinds of evidence. “[I]n practice,” Blackmun wrote, “a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations.” But this was the balance struck by Rules of Evidence, since they were “designed not for the exhaustive search for cosmic understanding but for the particularized resolution of legal disputes” (Daubert 1993, p. 597).

Both civil plaintiffs and civil defendants proclaimed victory following the decision (Greenhouse 1993).

Because Daubert changed the focus of the admissibility inquiry, it is not surprising that it should have a different impact than was felt under Frye (Groscup et al. 2002). Although research is ongoing regarding Daubert’s actual impact, it appears that the new rule will be both more permissive and more strict than the old rule, depending on the culture of the field from which the expertise comes (Faigman et al. 1994). Under Frye, the test was whether the field accepted the proffered knowledge, whereas under Daubert the test is whether the proffered knowledge is valid. Daubert should be expected to be more liberal for knowledge coming from rigorous fields in which adequate testing is part of the research tradition. In those fields, such as DNA profiling, robust discussion and debate might indicate lack of agreement despite the existence of sufficient data on which courts might permit expert opinion. In contrast, Daubert should be expected to be more conservative for knowledge coming from fields in which there is little or no research tradition of rigorous testing, e.g., psychoanalysis. In such fields, consensus replaces critical inquiry. Because Daubert asks what data support the proffered opinion, fields that operate by consensus rather than through data collection should be expected to have difficulty passing muster.

Overall, however, it has become clear that civil defendants have more to celebrate than their counterparts. Many lower courts immediately understood Daubert to be a stand against so-called junk science. On remand to the Ninth Circuit, Judge Alex Kozinski (Daubert v. Merrell Dow Pharmaceuticals, Inc. 1995) interpreted Daubert as creating a brave new world in which trial court judges donned white coats and were forced to play scientist. Indeed, in Daubert itself, Chief Justice Rehnquist had concurred in part to register his lack of confidence in federal judges becoming successful “amateur scientists” (pp. 600–601). In their new capacity as gatekeepers, judges have brought greater scrutiny to bear on proffered expertise.

The impact that this increased scrutiny has had in criminal cases remains a controversial and still unresolved matter. Many commentators speculated early on that the forensic sciences relied upon by prosecutors would fare poorly under Daubert (Risinger & Saks 1996, Saks 1998). Little of it is supported by quality research (Risinger et al. 2002). Moreover, prosecutors, even more so than plaintiffs, bear the burden of proof at trial. The courts, however, with minor exceptions, have so far resisted applying Daubert vigorously to prosecutorial expert testimony.
This reluctance has resulted in opinions, especially for example on the subject of fingerprinting, that stretch credulity beyond the breaking point (e.g., United States v. Havard 2000). Many legal scholars have commented on the seeming hypocrisy of applying Daubert strictly to civil plaintiffs and leniently to criminal prosecutors (Seton Hall Symposium 2003). This is also the subject of social science research, which appears to validate the observed trend (Groscup et al. 2002, Groscup & Penrod 2003).

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, carefully scrutinized the proffered expertise and, in some detail, pointed out its many deficiencies. Moreover, the Court revised Blackmun’s statement that the trial court’s obligations were restricted to methods and principles, and 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” (General Electric Co. v. Joiner, p. 146).

Two years later, in Kumho Tire Ltd. v. Carmichael (1999), the Court closed the biggest loophole left from Daubert. Justice Stephen Breyer, the heir apparent to Blackmun’s label as resident scientist, held that a court’s gatekeeping responsibilities extended to all expert opinion, not just that of the scientific variety. The Kumho Tire opinion is replete with statements regarding trial courts’ 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

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 plaintiffs’] 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

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 upon which virtually all Daubert-watchers, whether its supporters or detractors, would agree. 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. The rule does not contain degree requirements. An expert must only be qualified “by knowledge, skill, experience, training, or education.” The case law, therefore, is replete with statements of deference to proffered experts’ qualifications, recognizing the value of experience and practical training.

At the same time, the case law is also rich in demands that experts be better qualified than those that have been offered. 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. Examples of this might be a nonlicensed therapist seeking to testify regarding the competency of a defendant to stand trial or a research psychologist without an adequate background in statistics and research methods. 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 as well as 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, “[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. 2002). 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 situation similar to that of medical science, in that many practitioners are trained as generalists although there are a number of subdisciplines in which psychologists can specialize. On the whole, however, this issue appears to arise more frequently with medical doctors then it does with psychologists. One reason for this 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. In particular, many clinical psychologists define themselves by either theoretical orientation or the developmental stage of the subjects of study. These categories 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.

The more likely qualifications debate to be heard in courtrooms 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 on 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. 2002). 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.

Increasingly, courts tend to measure qualifications against the content of the proffered testimony. Consequently, an expert must demonstrate facility with the actual subject matter of the facts in dispute. Accordingly, a DNA technician cannot testify about the underlying science and a police officer is not permitted to testify on the intoxicating effects of blood alcohol level. However impressive may be their technical expertise, if expert witnesses do not have sufficient grounding in the underlying science, they should be precluded from testifying regarding the validity of the techniques they employ. The same holds true for psychological experts. Clinicians may have little familiarity with the extant research literature on the multitude of subjects that they put into practice. A clinician trained to apply
the *DSM-IV*, for instance, may have little idea what sort of research is available to validate the diagnoses he or she employs daily. Similarly, a research psychologist might be unqualified to conduct a diagnostic interview.

**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 must create the chain of inferences necessary to construct a case, every piece of evidence must at least be 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, proof that a defendant was in San Francisco the day a robbery occurred is relevant to the government’s case if the crime took place in San Francisco, but irrelevant if the crime were wire fraud that could have occurred during a phone call from anywhere. Similarly, expert testimony that a defendant lacks volitional control over his or her behavior is relevant in a state employing the American Law Institute test for insanity, but it is irrelevant in states using the M’Naghten 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 American Law Institute test for insanity, lack of volitional control is an element of the defense, but under M’Naghten it is not.

The notion of fit as a component of the relevancy inquiry involves a somewhat different level of analysis. As a general matter, it involves 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 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 (Monahan & Loftus 1982).

**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
seven years before. The three principal rules, Rules 701, 702, and 703, underwent revision. Rules 701 and 703 were both revised with the intent of shoring up identified weaknesses and ensuring the smooth operation of Rule 702.

Rule 701 was strengthened to make sure that opinion evidence that should be evaluated under Rule 702 did not slip in through the backdoor of the lay opinion rule. Rule 701 only applies to lay opinion testimony that was within the personal knowledge of the witness. The rule provides as follows:

If the witness is not testifying as an expert, the witness' testimony in the form of opinions or inferences is limited to those opinions or inferences which are (a) rationally based on the perception of the witness, (b) helpful to a clear understanding of the witness' testimony or the determination of a fact in issue, and (c) not based on scientific, technical, or other specialized knowledge within the scope of Rule 702.

Rule 703 was also strengthened and, in particular, the rule was changed to end the frequent practice of parties using experts to introduce otherwise inadmissible hearsay statements. The new rule 703 provides as follows:

The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence in order for the opinion or inference to be admitted. Facts or data that are otherwise inadmissible shall not be disclosed to the jury by the proponent of the opinion or inference unless the court determines that their probative value in assisting the jury to evaluate the expert’s opinion substantially outweighs their prejudicial effect.

The original Rule 703 suffered from an Achilles' heel that rendered it of great service to practitioners with hearsay evidence they were otherwise having difficulty getting admitted. Rule 703 appeared to allow experts to testify to hearsay statements that would otherwise be excluded, so long as they were “facts or data . . . of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject” (Carlson 1992). It thus operated as a conduit for otherwise inadmissible evidence. Rule 703 was amended to close this loophole. The Ninth Circuit observed in Turner v. Burlington Northern Santa Fe Railroad Co. (2003) that the new Rule 703 requires a court to ask two questions when evaluating otherwise inadmissible evidence:

The first question is “whether the facts are of a type reasonably relied on by experts in the particular field.” The second question is whether the probative value of the underlying data substantially outweighs its prejudicial effect.

Initially, as a simple factual matter, the court must determine whether the facts or data truly are of the type that an expert in the field would reasonably rely upon. Ordinarily, this will depend on the credibility of the expert’s warranting that this
is so, as well as what evidence is adduced by the opponent calling this matter into question. Underlying this determination, however, is the Daubert demand that judges not simply rely on professional fields to determine what sorts of facts or data might be admitted. Typically, courts consider the initial factual question of whether the facts or data are the sort that experts from that particular field would rely upon separately from the reliability of the field more generally (Montgomery Cty. v. Microvote 2003).

The second question requires a balancing of probative value and prejudicial effect. In Turner (2003), the Ninth Circuit emphasized that evidence that is otherwise “inadmissible will be kept out unless the court determines that any potential prejudice is substantially outweighed by the probative value” (p. 1062). This assessment lies within the discretion of the trial court and will not be disturbed on appeal unless it constitutes an abuse of discretion.

The revision to Rule 703 is likely to have a disproportionate impact on psychologists and medical doctors. These professionals are the most likely, given the nature of their respective disciplines, to relate statements made by clients and patients as a basis for the opinions they offer in court. These experts will no longer be permitted to testify to hearsay statements unless the statements would be admissible under a hearsay exception or exclusion or the statements’ probative value substantially outweighs their prejudicial effect.

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 (a) testing, (b) error rate, (c) peer review and publication, and (d) general acceptance. Many commentators expressed fear that Daubert might be used woodenly by unsophisticated courts. Indeed, some courts initially believed that the Daubert factors constituted a “checklist” and sought to apply them mechanically. From the start, however, the Court made clear that the Daubert factors were not exclusive, and this fact was later emphasized in Kumho Tire.

Moreover, courts and commentators have usefully suggested additional factors that might be employed to complete the assigned task (Faigman et al. 2002). More importantly still, 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, “[t]hough, 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).

Because psychological testimony is ostensibly based on scientific knowledge, the *Daubert* factors are a reasonable starting point in most cases. As the Court in *Certain Underwriters at Lloyd’s v. McDermott International, Inc.* (2002, p. 4) put it, “in the vast majority of cases, the district court should first consider the *Daubert* factors before addressing whether other factors are relevant to the particular case.” In *Daubert* (1993, p. 590), the Court emphasized that “[i]n a case involving scientific evidence, evidentiary reliability will be based on scientific validity.” We, therefore, use the *Daubert* factors below as a launching point to begin 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 have been adequately tested (Black et al. 1994). Ultimately, the challenge of *Daubert* lies in measuring this adequacy.

Early on, it was generally thought that the form in which testimony was offered, scientific or nonscientific, mattered. Several courts proceeded on the theory that scientific evidence was usefully distinguished from nonscientific evidence. This was based foremost on the belief that some sort of bright line could be drawn between the two. Second, many courts also believed that scientific evidence cast a wondrous aura that might overwhelm triers of fact. As a consequence, these courts maintained, scientific experts should have to overcome many more hurdles than “specialists” or “technicians” if they were to be permitted to speak to jurors.

The *Kumho Tire* (1999) Court rejected the distinction between science and nonscience and refocused the inquiry. Justice Breyer, writing for the Court, said that “it would prove difficult, if not impossible, for judges to administer evidentiary rules under which a gatekeeping obligation depended upon a distinction between ‘scientific’ knowledge and ‘technical’ or ‘other specialized’ knowledge. There is no clear line that divides the one from the others” (p. 148). The Court also rejected the proposition that jurors are likely to be more impressed with the science label than with technical or specialized expertise. Breyer observed that all experts are given considerable leeway to offer opinion evidence. Any concern with triers of fact yielding to the authority of scientific opinion was similarly present with all expert testimony.
The more difficult issue presented under *Daubert* involved the manner in which trial courts should carry out their gatekeeping duties given the multitude of disciplines that enter courtrooms every day. The *Daubert* factors could not be applied to every sort of expert, and the rigor represented by those factors might not be appropriate in all cases. The Court had outlined an approach by which real estate appraisals, accounting, DNA technology, social psychology, and clinical medicine would all have to be judged. *Kumho Tire* (1999) extended *Daubert* beyond just “scientific” expertise but elided over the question of how trial courts should measure the validity of the great variety of expertises that courts confront daily. The Court stated:

The conclusion, in our view, is that we can neither rule out, nor rule in, for all cases and for all time the applicability of the factors mentioned in *Daubert*, nor can we do so for subsets of cases categorized by category of expert or by kind of evidence. Too much depends upon the particular circumstances of the particular case at issue... [A] trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of the expert testimony (*Kumho Tire* 1999, pp. 150–152).

Properly understood, the *Daubert* factors represent a set of criteria by which fairly traditional scientific evidence might be judged. 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 the criteria to assess the success of that knowledge.

As regards the myriad of “subsets of cases categorized by category of expert or by kind of evidence,” other factors in addition to, or replacing, the *Daubert* set of four might be useful. This ought to be true whether the proffered expert is an art historian or a structural engineer. Professionals in different fields inevitably have criteria by which they measure the merit of the expert information that is produced by members of the field. If courts find a discipline’s evaluative criteria to be adequate, they can employ them just as the *Daubert* Court did in the category of expertise (i.e., epidemiology) involved in that case (Denbeaux & Risinger 2003). Increasingly, in order to help manage the many forms expert opinions take, scholars have sought to posit taxonomies of expertise to aid this effort (Gross & Mnookin 2003, Risinger 2002). The basic objective of this work is to devise factors and strategies by which a vast assortment of expertise might be evaluated. Courts have yet to embrace this fledgling movement, but they appear amenable to the effort. Some courts have even joined in. For example, the court in *Ebenhoech v. Koppers Industries, Inc.* (2002) suggested nine factors to be considered for engineering experts.

Psychologists have yet to embark on a similar taxonomic endeavor within their field, with most commentators seemingly willing to accept *Daubert*’s factors at
face value. Indeed, so far there has been a surprising quiet among psychologists who might have been expected to be critics of Daubert. Psychology presents a microcosm of modern expertise as it is found in courts today. But those psychologists with little data who nonetheless seek to testify have yet to complain very loudly about Daubert’s rigorous standards. Instead, these experts argue that the anecdotal method is a form of testing that meets Daubert. They also tend to rely on the criteria of peer review and publication and general acceptance to buttress their arguments for admissibility. We consider these issues below.

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 scientists substantial guidance. He merely stated, “[I]n the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error” (Daubert v. Merrell Dow Pharm., Inc. 1993, p. 594). He thus appeared to limit his concern with error to techniques like DNA profiling or polygraph tests that might have an identifiable error rate with repeated applications. Blackmun’s simplistic treatment of error is more perplexing still because the principal scientific subject of the Daubert case was epidemiology, a research area that employs the concept of error rate in complex and robust terms. Lower courts since Daubert have done little to remedy the deficiencies in Blackmun’s analysis of error rates. Indeed, their tendency is to gloss over or ignore this factor altogether. As any bright undergraduate psychology major knows, error in science is a fairly complex multidimensional concept. Yet courts hardly consider even the most rudimentary and unidimensional aspects of it. For instance, courts rarely distinguish between Type I and Type II error when they discuss this factor. Courts also have provided precious little discussion of random or systematic error, either as a general matter or in relation to specific proffered expert evidence.

When courts discuss error, they typically limit their analyses to methodological problems that might bias the results. These concerns might generally be better considered under the first Daubert factor concerning the adequacy of the testing. Courts usually limit their discussions to such things as debates about the generalizability of animal testing, especially in toxic tort cases, and the use of undergraduates in psychology experiments. This sort of error is mainly nonquantifiable, which may be one of its principal virtues as far as courts are concerned.

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). In addition, some work that appears outside the strict scientific peer review process is well done and well worth relying upon. 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.

A somewhat less noticed aspect of the peer review and publication factor is the more general dynamic that is present in healthy scientific communities. In describing the peer review criterion, Blackmun observed, “submission to the scrutiny of the scientific community is a component of ‘good science’ . . . because it increases the likelihood that substantive flaws in methodology will be detected” (Daubert 1993, p. 593). Peer review and publication is just one instance of a much more general phenomenon of critical self-assessment that mature sciences engage in as a matter of course.

Psychology, like many applied scientific fields, does not always carry out this checking function successfully. 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 (Faigman 1989).

There is an assortment of reasons for this phenomenon. Good applied psychological research tends to be expensive. Given limited resources, it is not surprising that many researchers would prefer to extend knowledge beyond the last study, rather than “waste” time, money, and energy replicating another’s research. In addition, prestige and professional recognition are associated with investigating new hypotheses rather than with confirming old ones. In the better research programs, researchers strive to replicate previous work in the process of extending those discoveries or moving the enterprise in new directions. The step-by-step construction of empirical knowledge, through the efforts of independent research groups, is a sign of a healthy field.

An even greater threat to the checking function inherent in good science is the concern that significant areas of psychology serve policy agendas, whether of the researchers or of certain political communities more generally (Redding 2001). Research that appears to substantiate a good policy outcome may be resistant to subsequent work. Later studies that cast doubt on politically fashionable research findings may not be a career-enhancing move. Although this concern might be leveled at most of the social sciences—and many of the applied physical sciences as well—psychologists with an interest in affecting legal discourse should be greatly concerned if their findings are too easily dismissed as the product of politics. Scientists with strong policy agendas are readily discounted by courts, however strong their research program might be.
General Acceptance  Like peer review and publication, the use of 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 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 STRUCTURE OF PSYCHOLOGICAL SCIENCE AND THE RULES OF ADMISSIONABILITY

Law and science approach the empirical world in fundamentally different ways. Whereas science attempts to discover the universals hiding among the particulars, trial courts attempt to discover the particulars hiding among the universals (Faigman 1999, p. 69). Research on eyewitness identification, for example, examines those factors that might interfere with memory and perception and thus could render eyewitness accounts less reliable. The result has been a robust list of general factors that, if present, may be relevant in particular cases (Wells & Olson 2003). The trier of fact, of course, must decide whether a particular witness’s testimony is reliable. In most cases, this means that, if allowed, experts provide triers of fact with general information and leave it to them to apply these insights to the particular case. Another well-regarded area of psychological study, work on the suggestibility of children’s memory, operates very similarly (Bruck & Ceci 1999). Research has produced substantial insights regarding children’s memory, but it does not speak directly to the question of whether a particular child-witness is accurate. In many contexts, however, experts seek to supply both the general knowledge and the specific application of that knowledge to the particular case. For instance, recitation of the general research basis for posttraumatic stress disorder (PTSD) is ordinarily followed by clinical testimony informing the trier of fact that the subject is or is not an example of the general phenomenon. This combination of general and specific is endemic to the law and science intersection and presents a substantial challenge to both lawyers and scientists.

Although not easily resolved, the phenomenon of the general nature of science and the particularized interest of the law is generally understood by lawyers and scientists. Courts so far have dealt with this issue primarily in the medical context in which it is starkly presented. There, courts routinely distinguish between “general causation” and “specific causation.” The former refers to whether, for example,
research demonstrates that a substance can cause cancer; the latter refers to whether the substance caused a cancer in a particular patient. Courts agree that general causation must be shown before evidence of specific causation is allowed.

Although somewhat less developed in the case law, psychology presents an essentially identical situation to that of medicine. The basic schema now being developed under Daubert parallels the schema suggested in a series of articles published in the 1980s (Monahan & Walker 1986; Walker & Monahan 1987, 1988) that identified three basic ways in which the law employs social science research. These are as social authority, social facts, and social frameworks. Social authority refers to social science research relevant to the determination of facts that are used in the formulation of legal rules or policy. According to this proposal, social authority is analogous to legal authority and should be consulted similarly. Hence, judges would consider social science “precedent” (i.e., past research) as presented through briefs, oral arguments, and sua sponte (of one’s own accord). Relevant research would then be incorporated into the judge’s conclusions of law. Alternatively, social science might be relevant to facts that are specific to the litigation. Most survey evidence commissioned for a given case fits into this category. Social fact evidence would be presented to the trier of fact through expert testimony. The final category is social framework evidence, in which social science research has relevance as a combination of social authority and social fact. In social frameworks, some issue in the particular dispute is claimed to be an instance of a social scientific finding or theory of general import. All three categories have been affected by Daubert and its progeny, though social framework evidence supplies the lion’s share of the case law.

Social Authority

Although the Daubert decision concerns the meaning of rules of evidence, and most of the subsequent case law involves the practical application of the Daubert test to expert evidence at trial, its significance extends well beyond the four corners of the courtroom. Empirical judgments enter legal analyses in ways well outside of the ordinary evidentiary process. In particular, empirical judgments are often an essential component of the policy-making function of courts, especially in the context of constitutional law. In abortion cases, for example, the empirical judgment of when a fetus becomes medically viable sets the time at which states can prohibit abortions (Planned Parenthood v. Casey 1992). Other empirical judgments of import to abortion litigation are of a more psychological nature, including whether previability abortion regulations, such as spousal notification or 24-hour waiting provisions, constitute a “substantial obstacle” to the exercise of a fundamental right (Casey 1992). In capital sentencing cases, empirical judgments concerning the psychological deterrent effect of execution, its possible discriminatory application, and whether juries from which opponents of capital punishment have been removed are biased against the defendant have been the subjects of the Supreme Court’s attention (Gregg v. Georgia 1976, McCleskey v. Kemp 1987, and
Lockhart v. McCree 1986, respectively). Additional examples of the Supreme Court’s confronting empirical judgments upon which psychological research was brought to bear include school segregation (Brown v. Board of Education 1954), the testimony of child witnesses via closed-circuit television in sexual abuse cases (Maryland v. Craig 1990), the consequences of permitting physician-assisted suicide (Washington v. Glucksberg 1997), the effects of jury size (Ballew v. Georgia 1978), and the learning styles of men and women (United States v. Virginia 1999), among many others (Faigman 2004).

The relationship between Daubert and empirical judgments relevant to the making of constitutional law is complex and courts and scholars have yet to fully work through the subject’s many complexities. Consider, for example, the empirical judgments called for in the litigation involving the Violence Against Women Act (42 U.S.C. §13,981), which was struck down by the Supreme Court in United States v. Morrison (2000) on the basis that Congress did not have the authority to pass the Act under the Commerce Clause. Congress had held extensive hearings on the question of whether domestic violence substantially affected interstate commerce, including considering testimony on the psychological consequences of such violence. Congress explicitly found that violence against women was a national problem that substantially affected commerce and thus the Constitution authorized federal regulation. Opponents challenged this factual finding, arguing that regulation of domestic violence was within the purview of the states. The Supreme Court agreed with this argument and invalidated the statute. Initially, however, this factual question had to be litigated before a trial court. For a trial court hearing this matter, at least three sources of evidence might be available to consider. The first is expert testimony proffered at trial by the parties, the second is the congressional record of the hearings upon which the law was based, and the third is briefs filed by the parties and amici that substantively discuss the science. The lessons of Daubert apply differently to these three sources of evidence.

Expert testimony regarding constitutionally relevant empirical judgments introduced at trial are subject to the procedural strictures of Daubert and its progeny (National Abortion Federation v. Ashcroft 2004). Accordingly, expert opinion is not admissible unless the proponent of the evidence demonstrates that it is based on reliable and valid methods and principles and that the opinion reasonably follows from the underlying research. In contrast, because congressional hearings and briefs do not come in through ordinary evidentiary procedures, Daubert and the Rules of Evidence do not formally apply to this kind of empirical information. This is not to say, however, that Daubert or the principles it represents should not be applied to this kind of empirical information. Although legislative hearing testimony and briefs are not subject to admissibility determinations, courts must nonetheless weigh the information presented at the hearings or contained in the briefs in making the ultimate constitutional determination. Daubert, and in particular its principle of scientific sophistication, can be employed usefully to evaluate the value of extrajudicial expert opinion. Therefore, courts should consider, for instance, whether the research presented in the congressional testimony or that is
set forth in a brief, was adequately tested, had reasonable error rates, was subject
to peer review and publication, and is generally accepted among knowledgeable
members of the field. To the extent that psychological or other research found in
legislative hearings or briefs fails a Daubert-styled review, courts should give it
less weight or no weight at all.

Social Facts

Psychological researchers sometimes conduct empirical studies to determine a
specific “social fact” at issue at trial. This most often occurs in trademark litigation.
The Trademark Act of 1988 states that the federal Patent and Trademark Office
will refuse to register a new trademark if it so resembles a trademark already
registered to another person “as to be likely . . . to cause confusion.” Psychological
surveys—or, sometimes, true experiments—to establish whether consumers or
potential consumers are likely to be confused, or are actually confused, between
products with similar trademarks have been admissible in court at least since
Zippo Manufacturing Company v. Rogers Imports, Inc. in 1963. Increasingly, the
admissibility of expert psychological testimony on consumer confusion is being
held to Daubert standards. For example, in Bicardi v. New York Lighter (2000), the
Court applied Daubert to a survey proffered by the plaintiff, the manufacturer of
Bicardi rum, addressing whether consumers would be confused by the defendant’s
naming its cigarette lighter “Bicarbi.” Excellent guides now exist (Diamond 2002)
that walk psychologists through the steps necessary to conduct and analyze survey
research in trademark and related cases that will survive Daubert review.

Obscenity is a second area that in the past has often seen psychologists testify
as “social fact” witnesses. The Supreme Court, in Miller v. California (1973),
held that one of the prongs of the test to determine whether a publication or film
was obscene was “whether the average person, applying contemporary commu-
nity standards, would find that the work, taken as a whole, appeals to the prurient
interest.” Surprisingly, however, no published case has yet subjected a psycholog-
ical survey of community standards regarding obscenity to a Daubert analysis.
The likely reason for this nonapplication of Daubert is that obscenity prosecutions
declined to almost zero during the Clinton administration (Blum 2004), and so few
surveys of community standards were proffered in court. The Bush administration,
however, has vowed to begin the vigorous prosecution of obscenity cases, so the
first applications of Daubert to the issue of community standards can be expected
soon.

When those applications take place, courts will have to resolve an issue of op-
erational definition on which existing case law is conflicted: Are the “community
standards” at issue in obscenity prosecutions to be determined by assessing and
aggregating the personal standards of a representative sample of the relevant com-
munity, or by asking respondents about their perceptions of the standards of others
in the community? Linz et al. (1991), in the context of an actual obscenity case,
demonstrated how assessing and aggregating the personal standards of community
members leads to much more acceptance of pornography than asking community members their perceptions of their neighbors’ standards. As in Lake Wobegon, everyone appears to be more tolerant than average. Given these results, they recommend that courts determine “community standards” by expert testimony that uses the former rather than the latter methodology.

The great bulk of what is usually seen as psychological “fact” testimony is presented by clinical psychologists rather than by researchers who conduct original studies that might be introduced in court (Slobogin 2002). Judging from the content of forensic handbooks such as Melton et al. (1997), Heilbrun et al. (2002), Goldstein (2003), and Grisso (2003), the factual issues that clinical psychologists most often address in court include a variety of legal competencies (e.g., to stand trial, to plead guilty, to testify, to waive rights, to be executed, to be a parent, to consent to or to refuse treatment, to make a will), the existence of mental disorder for various legal purposes (e.g., civil commitment, the insanity defense and diminished capacity, workers compensation, tort damages, invocation of the Americans With Disabilities Act), and predictions of specified kinds of future behavior—often, violence—in many legal contexts (e.g., civil commitment, sex predator commitment, criminal sentencing, child custody). Daubert places the onus squarely on the clinician-witness to demonstrate the empirical validity of the factual conclusions to be offered in court. There is, of course, a voluminous research literature underlying many (though not all) of the mental disorders listed in the DSM-IV. Recent years also have seen the development of a wide variety of what Grisso (2003, p. 42) has called “forensic assessment instruments.” Some (though, again, by no means all) of these forensic assessment instruments are supported by an extensive body of tested, published, and peer-reviewed validation research and are, or rapidly are becoming, generally accepted in the field of forensic evaluation. For example, Grisso & Appelbaum (1998) have developed and validated an instrument to assess competence to consent to treatment, and Poythress et al. (2002) have developed and validated an instrument to assess competence in the criminal adjudication process, derived both from an analysis of the underlying legal concerns, as revealed in the case law, and from basic psychological research in decision making. Such instruments should have little difficulty in surviving Daubert review (e.g., Commonwealth v. Morasse 2001).

Expert opinion on social facts bears some resemblance to the social framework evidence discussed in the next section. Opinions about competence, mental disorder, or future violence, contain both a general component and a specific component. For instance, “dangerousness” determinations should depend on well-validated measures that are appropriate to use in the particular forensic context in which they are employed. Although clinical assessments of future violence may be valid and reliable when based upon “general” actuarial measures, those assessments may be invalid if based solely on unstructured clinical judgment limited to a “specific” patient. Similarly, a test that is employed for therapeutic purposes may not be valid for forensic purposes. For example, most courts agree that while “penile plethysmographs can help in the treatment and monitoring of sex offenders,
these tests have no indicia of reliability as evidence at trial” (Doe v. Glanzer 2000, p. 1266).

Social Frameworks

If courts treat social frameworks such as the “battered woman syndrome” or the “rape trauma syndrome” the way they treat what might be called medical frameworks, they should require that both general causation and specific causation be demonstrated independently under Daubert. Moreover, without adequate proof of general causation, no testimony at all on specific causation will be permitted. Even with proof of general causation, proof must still be forthcoming that specific causation is within the capabilities of the science and the proffered expert. Much of psychology is likely to struggle to meet these standards, as medicine has struggled as well (Kassirer & Cecil 2002). As it turns out, courts themselves have struggled to devise a consistent response to social framework evidence and there is much variability in the case law (Monahan & Walker 2002). Indeed, U.S. Courts of Appeal disagree on the amount of rigorous proof that is necessary to permit expert opinion on questions of medical causation. The Third Circuit, for instance, has adopted a relatively lenient approach under Daubert and has allowed a good deal of clinical judgment to support the opinion (Heller v. Shaw 1999). The Fifth Circuit, in contrast, has taken a decidedly more rigorous approach and only allows opinion on medical causation that is supported by substantial research (Black v. Food Lion, Inc. 1999). The Supreme Court will likely have to settle this issue before long. However the courts eventually resolve these differences will have a substantial impact on psychology and the experts who testify on the subject (Fiske & Borgida 1999).

Clinical assessments in psychology are closely analogous to clinical assessments in medicine. As noted above, the initial question presented in medical cases is whether substantial research exists to support an inference of general causation. In Daubert, for example, the issue was whether Bendectin, a morning sickness drug, causes birth defects. Research on general causation took a variety of forms, including epidemiology. The key issue was whether the exposed population had a higher incidence of birth defects than the nonexposed population. In the end, the best research indicated no statistically significant increased risk of birth defects in the population exposed to Bendectin (Sanders 1993).

Consider PTSD as it is used in “rape trauma syndrome” sexual assault cases. It is quite analogous to the Bendectin case. One should expect good quality research demonstrating a higher incidence of PTSD in the affected population than in the general population. Research that fails to use a comparison group, therefore, has little probative force. In addition, courts should expect that good research methods would be used to assess the relative risk, just as would be true in epidemiological work. To date, however, studies on rape and the short-term development of PTSD have largely failed to use comparison groups or have suffered from other methodological flaws (Frazier 2002).
With psychology, as with medicine, courts sometimes muddle the issues of general and specific causation. Consider, for example, the subject of repressed memories. In Logerquist v. McVey (2000), the Arizona Supreme Court believed that proffered expert testimony regarding the plaintiff’s claim of repressed memories of abuse committed by the defendant was not “scientific” and thus did not have to pass muster under the Frye test. In effect, the Logerquist Court found that there was no general causation issue by which a specific causation deduction had to be ascertained. The only issue was specific causation, which the Court believed could be the product of inductive reasoning. The Court explained that Frye “is applicable 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.” This bypass of Frye (or Daubert) is called the opinion rule, because it allows an expert to offer his or her opinion on technical subjects without clearing admissibility hurdles so long as it is based on personal experience. Several states employ the opinion rule, including California and Florida. The opinion rule, to put it mildly, represents approximately a sixteenth-century understanding of the scientific method.

An expert’s statement that a person had a repressed memory that has now been recalled is a product (must be a product) of both general causation and specific causation inferences. However implicit the expert wishes to leave it, the statement that person X had repressed memories is actually saying that person X is an instance of a more general set of similarly situated people who have, or have had, repressed memories. The Court sought to permit testimony on the particulars without any proof of the universals. According to the Logerquist view of the scientific method, an expert should be allowed to categorize an observed phenomenon, such as repressed memories, without first having to show that such a category exists. What the Court called inductive reasoning is simply hypothesis formation through anecdotal experience, but without hypothesis testing through formal scientific inquiry. The plaintiff’s expert 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 his “experience” give him this information?

Contrary to Logerquist, the key to success as an expert should turn out to be the same as the key to success in science—substantial amounts of high-quality research on general causation. Such work not only permits experts to rule in certain factors as possible causative agents, but also to rule out seemingly plausible alternative factors. Much contemporary psychology, as much contemporary medicine, will fail to meet this standard. This is so for many clinical tests relied upon regularly, such as the Rorschach or other projective techniques (Garb et al. 2003). It is also true for many of the myriad syndromes that seem to replicate at extraordinary rates [e.g., “shopping addiction” (United States v. Roach 2002)].
As to specific causation, even some of the best work in psychology provides little basis for expert opinion on this topic. Consider the premier example of good psychology and law over the past several decades, research on the reliability of eyewitness identifications (e.g., Wells & Loftus 2003). This research has done an excellent job of identifying a wide variety of factors that might interfere with eyewitness perception, memory, and recall. Researchers have demonstrated rather less interest in factors that might aid eyewitness accuracy. Yet, there is no research whatsoever on whether experts are able to distinguish reliable from unreliable eyewitnesses. Instead, researchers have studied the phenomenon at a general level and, with only an occasional exception, researchers do not attempt to offer opinions on the reliability of specific eyewitnesses.

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 2003). 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. 2002). Cecil & Willging (1994) found that judges relied little on this appointment power; 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 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
In federal cases, court experts assume primarily one of two forms. The first, properly termed court-appointed experts, are selected mainly to assist jurors in determining 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

It is sometimes suggested, especially by legal scholars and judges, that the Daubert standard should be relaxed in light of the enormous difficulty in studying human behavior (Kinports 2004). Raeder (1996), for example, has questioned the application of strict standards to a difficult-to-test subject like domestic violence. Obviously, such topics cannot be studied experimentally. While domestic violence or the effects of severe trauma are surely challenging subjects for psychologists to research, in principle they present no greater obstacles than many other subjects of interest to the law, such as DNA typing, engineering, and epidemiology.

In this context, we note that our review appears 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 et al. (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 (2002) 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 (2003), both of whom have been recent presidents of the American Psychological Association. Fox (2000, 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 (2003) 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.” He believes that Lilienfeld and others have “gone overboard in their enthusiasm for scientific vetting of therapeutic techniques” (Goode 2004). Such vetting, Levant believes, has already hurt practitioners by restricting the interventions that are reimbursed by insurance companies, and “could create additional hazards for practitioners in the courtroom if empirically validated treatments are held up as the standard of care in our field” (Levant 2003).

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.” Neither is admitted. Whatever clinical value the unvalidated psychological assessment or treatment techniques may or may not have for patients, Daubert makes plain that testimony employing such techniques has no place on the witness stand. For psychologists, Daubert stands as a guardian at the courtroom door, warmly welcoming those with relevant and valid scientific evidence, and brusquely sending others on their way.

ACKNOWLEDGMENT

The authors would like to thank Shari Diamond, Kirk Heilbrun, Michael Saks, and Larry Walker for their comments and suggestions on a draft of this chapter.

The Annual Review of Psychology is online at http://psych.annualreviews.org

LITERATURE CITED

Black v. Food Lion, 171 F.3d 308 (5th Cir. 1999)
Blum V. 2004. D.O.J’s new porn police. Legal Times, Mar. 8, p. 1


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


Faigman DL, Porter E, Saks MJ. 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 Rev. 15:1799–835


Goode E. 2004. Defying psychiatric wisdom, these skeptics say “Prove it.” N. Y. Times, March 9, p. F1


Heller v. Shaw, 167 F.3d 146 (3d Cir. 1999)


Linerqvist v. McVey, 1 P.3d 113 (2000)

Maryland v. Craig, 497 U.S. 836 (1990)


Miller v. California, 413 U.S. 15 (1973)


Montgomery County v. Microvote Corp., 320 F.3d 440 (3d Cir. 2003)


Seton Hall Symposium. 2003. Expert admissibility symposium: What is the question to which standards of reliability are to be applied? Seton Hall Law Rev. 34:1–388


Turner v. Burlington Northern Santa Fe Railroad Co., 338 F.3d 1058 (9th Cir. 2003)


United States v. Roach, 296 F.3d 565 (7th Cir. 2002)

United States v. Roldan-Zapata, 916 F.2d 795 (2d Cir. 1990)


CONTENTS

Frontispiece—Richard F. Thompson xviii

PREFATORY
   In Search of Memory Traces, Richard F. Thompson 1

DECISION MAKING
   Indeterminacy in Brain and Behavior, Paul W. Glimcher 25

BRAIN IMAGING/COGNITIVE NEUROSCIENCE
   Models of Brain Function in Neuroimaging, Karl J. Friston 57

MUSIC PERCEPTION
   Brain Organization for Music Processing, Isabelle Peretz and Robert J. Zatorre 89

SOMESTHETIC AND VESTIBULAR SENSES
   Vestibular, Proprioceptive, and Haptic Contributions to Spatial Orientation, James R. Lackner and Paul DiZio 115

CONCEPTS AND CATEGORIES
   Human Category Learning, F. Gregory Ashby and W. Todd Maddox 149

ANIMAL LEARNING AND BEHAVIOR: CLASSICAL
   Pavlovian Conditioning: A Functional Perspective, Michael Domjan 179

NEUROSCIENCE OF LEARNING
   The Neuroscience of Mammalian Associative Learning, Michael S. Fanselow and Andrew M. Poulos 207

HUMAN DEVELOPMENT: EMOTIONAL, SOCIAL, AND PERSONALITY
   Behavioral Inhibition: Linking Biology and Behavior Within a Developmental Framework, Nathan A. Fox, Heather A. Henderson, Peter J. Marshall, Kate E. Nichols, and Melissa A. Ghera 235

BIOLOGICAL AND GENETIC PROCESSES IN DEVELOPMENT
   Human Development: Biological and Genetic Processes, Irving I. Gottesman and Daniel R. Hanson 263
CONTENTS

SPECIAL TOPICS IN PSYCHOPATHOLOGY

The Psychology and Neurobiology of Suicidal Behavior,
Thomas E. Joiner Jr., Jessica S. Brown, and LaRicka R. Wingate
287

DISORDERS OF CHILDHOOD

Autism in Infancy and Early Childhood, Fred Volkmar,
Kasia Chawarska, and Ami Klin
315

CHILD/FAMILY THERAPY

Youth Psychotherapy Outcome Research: A Review and Critique
of the Evidence Base, John R. Weisz, Amanda Jensen Doss,
and Kristin M. Hawley
337

ALTRUISM AND AGGRESSION

Prosocial Behavior: Multilevel Perspectives, Louis A. Penner,
John F. Dovidio, Jane A. Piliavin, and David A. Schroeder
365

INTERGROUP RELATIONS, STIGMA, STEREOTYPING,
PREJUDICE, DISCRIMINATION

The Social Psychology of Stigma, Brenda Major
and Laurie T. O’Brien
393

PERSONALITY PROCESSES

Personality Architecture: Within-Person Structures and Processes,
Daniel Cervone
423

PERSONALITY DEVELOPMENT: STABILITY AND CHANGE

Personality Development: Stability and Change, Avshalom Caspi,
Brent W. Roberts, and Rebecca L. Shiner
453

WORK MOTIVATION

Work Motivation Theory and Research at the Dawn of the Twenty-First
Century, Gary P. Latham and Craig C. Pinder
485

GROUPS AND TEAMS

Teams in Organizations: From Input-Process-Output Models to IMOI
Models, Daniel R. Ilgen, John R. Hollenbeck, Michael Johnson,
and Dustin Jundt
517

LEADERSHIP

Presidential Leadership, George R. Goethals
545

PERSONNEL EVALUATION AND COMPENSATION

Personnel Psychology: Performance Evaluation and Pay for Performance,
Sara L. Rynes, Barry Gerhart, and Laura Parks
571
CONTENTS

PSYCHOPHYSIOLOGICAL DISORDERS AND PSYCHOLOGICAL EFFECTS ON MEDICAL DISORDERS
Psychological Approaches to Understanding and Treating Disease-Related Pain, Francis J. Keefe, Amy P. Abernethy, and Lisa C. Campbell 601

TIMELY TOPIC
Psychological Evidence at the Dawn of the Law’s Scientific Age, David L. Faigman and John Monahan 631

INDEXES
Subject Index 661
Cumulative Index of Contributing Authors, Volumes 46–56 695
Cumulative Index of Chapter Titles, Volumes 46–56 700

ERRATA
An online log of corrections to Annual Review of Psychology chapters may be found at http://psych.annualreviews.org/errata.shtml