2009

Daubert v. Merrell Dow Pharmaceuticals and the Local Construction of Reliability

Robert R Robinson

Available at: https://works.bepress.com/robert_robinson/1/
*Daubert v. Merrell Dow Pharmaceuticals* and the Local Construction of Reliability

*Robert Robinson*

**Abstract**

Scholars considering how expert testimony will fare under *Daubert* often apply the four dicta referenced by Justice Blackmun (testing, peer-review, error rate, and general acceptance) to determine whether such testimony will be admissible. In this article I critique this approach, contending that admissibility decisions cannot be adequately predicted by *Daubert* itself. *Daubert* has no clear legal rule for judges to apply, has no cognizable position on the degree of scrutiny expert testimony should face, and has no clear stance—even given the dicta—on what constitutes “good science.” When combined with the relative autonomy trial judges possess in making admissibility decisions, *Daubert*’s essential ambiguity leads to what I call “local constructions of reliability,” disparate and often competing conceptions of what constitutes reliable expert evidence. What is considered reliable in one area of expert testimony, such as medical causation, will be quite different from what is required for another, such as handwriting expertise.

If *Daubert* leads to such variation among different spheres of expert testimony, how can we generalize or predict judicial decision-making? I argue that admissibility decisions can and should be modeled empirically. Viewing judges as goal-oriented actors, I transform likely goals for judges making *Daubert* decisions into relevant variables for empirical modeling. These goals include attention to the quality of expert testimony (if not always under the *Daubert* dicta), maintaining institutional stability and judicial autonomy when faced with controversial scientific claims, and advancing judicial policy preferences. I intend this discussion to serve as a template for further empirical work on *Daubert*. 
## Table of Contents

INTRODUCTION 1

ADMISSIBILITY “REGIMES”: A THEORETICAL FRAMEWORK 4
   Epistemology and Judicial Capacity in Assessing Expert Testimony 4
   The Federal Standard: Daubert v. Merrell Dow Pharmaceuticals 7
   Daubert and the Theoretical Framework: Realist, Constructivist, or Neither? 12

THE EMPTY VESSEL: DAUBERT AND THE LOCAL CONSTRUCTION OF RELIABILITY 13
   Law, Legal Realism, and Judicial Decision-Making 13
   The Institutional Context of Daubert Decisions 16
   Daubert’s Fractured Stance on Admissibility 16
      Admissibility in the Daubert Opinion 16
      Contrasting Admissibility Standards in United States v. Crisp 20
   Daubert’s Fractured Stance on Epistemology 22

EMPIRICALLY MODELING DAUBERT DECISIONS 25
   Judges as Goal-Oriented Actors 25
   Legal Goals: Assessing Methodological Quality 25
   Institutional Goals: Weighing the Effects of Admissibility Decisions on Courts 27
   Attitudinal Goals: Advancing Judicial Policy Preferences 30
   External Variables 31

CONCLUSION: NOTES ON GETTING DAUBERT DATA 33
Daubert v. Merrell Dow Pharmaceuticals and the Local Construction of Reliability

Robert Robinson*

I. INTRODUCTION

Over the next few decades, it is conceivable that scientists will be able to “see” the genesis of a thought. The development of functional magnetic resonance imaging, or fMRI for short, permits a visual approximation of the brain at work, enabling scientists to determine more exactly the physiological starting places of rational thought, emotions, and deception. While much of the attention given to the “new neuroscience” will focus on age-old problems of free will and determinism,1 fMRI’s potential to detect lies may be of more immediate and practical importance to the justice system. Given the centrality of credibility and truth-telling to court proceedings, it is unsurprising that neurologists are already examining if fMRI scans can provide valid and reliable evidence that a particular individual is lying.2

Expert testimony linking brain scans to deception will only reach federal juries if it meets the evidentiary standard of reliability set in Daubert v. Merrell Dow Pharmaceuticals.3 A practitioner or legal scholar interested in the use of fMRI scans in federal courts (and state courts that apply a Daubert-like test) must therefore consider how the current research holds up to this standard. At present, at least three such considerations have been published by legal scholars.4 The central claim of these

---

3 Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993). The reaction of courts to new technologies has been decidedly mixed. In the forensic context, the criminal justice system has eagerly and perhaps uncritically accepted technological advances that individuate criminal suspects, such as fingerprinting or DNA typing. Jennifer Mnookin, Fingerprint Evidence in an Age of DNA Profiling, 67 BROOK. L. REV. 13 (2001). Advances that attempt to assess the credibility of individual statements, however, have often been received with suspicion and even disdain. Ric Simmons, Conquering the Province of the Jury: Expert Testimony and the Professionalization of Fact-Finding, 74 U. CIN. L. REV. 1013 (2006).
4 Archie Alexander, Functional Magnetic Resonance Imaging Lie Detection: Is a “Brainstorm” Heading Toward the “Gatekeeper”?, 7 HOUS. J. HEALTH L. & POL’Y 1 (2006); Charles N. W. Keckler, Cross-Examining the Brain: A Legal Analysis of Neural Imaging for Credibility Impeachment, 57 HASTINGS L.J.
articles, after providing a layman’s summary of the fMRI technique and the research regarding its potential as a lie detector, examines whether current research meets the testing, peer-review, error rate, and general acceptance dicta that often comprise the de facto test for applying Daubert. The authors generally agree that while the detection of deception using fMRI may someday be admissible, it currently fails to pass muster under these criteria.

Given that federal judges often structure their admissibility opinions along similar lines, it is quite reasonable for the authors of these articles to take this approach to predicting judicial behavior. That said, I argue in this article that such an approach is largely mistaken, and that better descriptive understandings of how Daubert is applied, as well as how better predictions of how Daubert will be applied in the future, require empirical modeling. The problem lies not in the authors’ lack of knowledge regarding the underlying science (which they understand well), or their particular applications of the Daubert dicta (which are reasonable). The problem lies instead in the assumption that the Daubert precedent provides a global definition of reliability that can be extrapolated to specific areas of expert testimony.

I state simply that it does not. Law is most effective in guiding judicial behavior when the law has a relatively clear rule, a relatively clear substantive meaning, or where judges face meaningful appellate oversight. Daubert decisions fit none of these criteria. As a decision rule, Daubert does not dictate how an admissibility determination should be carried out. The opinion easily supports contrasting views on the appropriate stringency of the courts towards admitting borderline expert testimony; it also easily supports contrasting views on the epistemology of science that judges should adopt. Even the Daubert dicta, which suggest that expert testimony should be tested, peer-reviewed, generally accepted, and present error rates, are terms lacking a consensus definition or application. As precedent, Daubert is much like the inscription on the Mirror of Erised from Harry Potter and the Sorcerer’s Stone: “I show not your face but your heart’s desire.”

A fractured precedent with no central meaning, Daubert leads to what I call “local constructions of reliability”; what constitutes reliable expert testimony in one sphere of science may be quite different from what passes Daubert in another. For example, Daubert’s application during a summary judgment motion in a “toxic tort” may be quite different from a motion where criminal defendants seek to exclude the state’s forensic testimony. The growing Daubert literature finds this split to be real, finding some evidence that admissibility standards have tightened in civil cases while remaining

---


7 Margaret A. Berger, What Has a Decade of Daubert Wrought?, 95 AM. J. PUB. HEALTH 59 (2005); Lloyd Dixon & Brian Gill, Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision, 8 PSYCHOL. PUB. POL’Y. & L. 251, 261-78 (2002); Ronald L. Melnick, A Daubert Motion: A Legal Strategy to Exclude Essential Scientific Evidence in Toxic Tort Litigation, 95 AM. J. PUB. HEALTH 30 (2005). I leave for others arguments over whether the impact in civil cases is the result of legal constraints imposed by Daubert, the increased attention to expert testimony that Daubert may have provided, or because of external changes, such as an increasingly conservative federal judiciary, that took
unchanged in criminal ones. My point is a broader one: since Daubert neither constrains nor guides judicial behavior, we must look beyond doctrinal analysis to understand what takes place when expert testimony is on trial.

If I am correct, Daubert scholarship will greatly benefit from the creation of empirical models of judicial behavior. While Daubert does not lack for scholarly attention, empirical research on Daubert, whether descriptive or predictive, has been quite limited in relation to the whole. In a recent survey of work on Daubert, Saks and Faigman find only seven empirical studies of admissibility patterns and the factors that might influence judicial decision-making. Much of the legal literature regarding admissibility is geared towards normative discussions about the philosophy of science, institutional arguments about the courts’ ability to assess expert testimony, and the abstract merits and deficiencies of particular admissibility regimes. Though undoubtedly important, this array of arguments would be well complemented by widespread empirical analysis of Daubert decisions at the trial court level, as well as by studies which attempt to assess the impact and significance of particular variables—beyond the dicta—that might influence the outcomes of admissibility decisions.

The outline of this argument is as follows. In Part II, I briefly examine the general problems of epistemology and judicial capacity any admissibility regime must face, and trace how Daubert skirts taking a clear stance on either issue. In Part III, I develop my argument that Daubert leads to local constructions of reliability rather than establishing a global one. Lacking effective institutional oversight, a formal decision rule, or a shared understanding of either the stringency of the admissibility standard or the proper epistemology behind it, Daubert decisions take place in a position of considerable judicial autonomy. Accepting that judges may be guided or constrained by local, on-point precedent for specific areas of expert testimony, we still need explanations for how local constructions of reliability develop in the first place. When assessing the reliability of new types of expertise such as using fMRI to detect lies, these explanations become vital.

In Part IV, I survey the political science, empirical legal studies, and sociology of science literatures to glean potential variables for a decision-model of admissibility decisions. Judges undoubtedly care about the “reliability” of expert testimony, but their modes of analyses are as likely to be intuitive as deliberate, assessing reliability through the identity of the litigant or the reaction of professional and academics interest groups. I


9 A simple keyword search of “Daubert” of law review articles in LEXIS returns over three thousand hits.


also consider the impact of long-term concerns over the impact of Daubert decisions on court institutions, such as increased caseload, or how some forms of expert testimony may face special burdens to admission because they seemingly challenge the empirical foundations of the courts themselves. For example, in a few appellate-level Daubert decisions regarding the defendants’ attempts to introduce eyewitness expert testimony, the judges worried that such testimony might threaten the right of the jury to assess credibility.\(^{12}\) Finally, Part V concludes the article by briefly noting an important roadblock for future empirical research on Daubert, namely the collection of representative data.

II. ADMISSIBILITY “REGIMES”: A THEORETICAL FRAMEWORK

A. Epistemology and Judicial Capacity in Assessing Expert Testimony

American legal jurisdictions are paternalistic when it comes to the admission of evidence, expert or otherwise.\(^{13}\) Neither truth nor fairness are best served, our legal system holds, by a laissez-faire presentation of any evidence various litigants might wish to submit. Judges instead employ a ponderous set of filters and rules, in hopes of shielding the decision-making from cognitive biases that jurors may not know they have.\(^{14}\)

This gate-keeping process becomes dramatically more difficult for judges when expert testimony is assessed. The justification for judicial gate-keeping of expert testimony is similar to the general one, namely that juries lack the scientific training and awareness of their own perceptual and cognitive errors to separate bad science from good. Expert testimony, however, disrupts the standard assumption of special judicial competence, as judges may be no better equipped than juries in determining methodological quality. This well-known dictum, often attributed to Learned Hand, appears insoluble: if the legal system could adequately discern the validity of expert opinions, it should not need that testimony in the first place.\(^{15}\) Given this paradox, any legal regime designed to assess expert testimony will be imperfect. Absent the wholesale admission of all proffered expertise, however, some test must be adopted.

The choice of potential admissibility regimes from alternatives such as the “market test,”\(^{16}\) the general acceptance test from Frye v. United States,\(^{17}\) or Daubert

\(^{12}\) United States v. Hall, 165 F. 3d 1095, 1107 (7th Cir. 1999); United States v. Amador-Galvan, 1997 U.S. App. LEXIS 5320 at *4-5 (9th Cir. March 14, 1997); United States v. Brien, 59 F. 3d 274, 276-78 (1st Cir. 1995).

\(^{13}\) Joseph Sanders, Kumho and ‘How We Know’, 64 LAW & CONTEMP. PROBS. 373, 406-07 (2001).

\(^{14}\) This central assumption regarding the degree of juror biases—at least relative to the biases of judges—has come under question; see, e.g., Luke M. Froeb & Bruce H. Kobayashi, Native, Biased, Yet Bayesian: Can Juries Interpret Selectively Produced Evidence?, 12 J. LAW. ECON. ORGAN. 257 (1996).

\(^{15}\) Learned Hand, Historical and Practical Considerations Regarding Expert Testimony, 15 HARV. L. REV. 40, 54 (1901). Hand’s concern dealt primarily with juries, but the problem would apply to any trier of fact not versed in the substantive field of expertise at hand.

\(^{16}\) This regime permitted expert opinions to be admitted in court if the opinion had been validated by the market in some way, whether because others had paid that expert in private endeavors, or because that expert claim engendered a product or service consumers had purchased. David L. Faigman et. al., Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence, 15 CARDOZO L. REV. 1799, 1803-06 (1994).
depend in large part on two criteria: the nature of “science” the legal system envisions, and the capability of judges to determine whether the evidence meets the chosen criteria. Selecting an epistemology to ground an admissibility regime is a difficult task, exacerbated by the sheer range of argument regarding whether scientific findings are “real.” At one end of this spectrum, philosophers such as Comte have offered an almost a naïve faith in scientists to deduce the nature of the universe from methods that transcend all problems of culture, ideology, and practicality. At the other end, epistemological anarchists offer the proposition that science bears little if any connection to reality, its findings instead wholly constructed by society. Between these poles of positivism and anarchism lie two more feasible epistemologies for courts to adopt: realism and constructivism. Realists believe that scientists may reveal an underlying reality through the scientific method, at least in probabilistic terms. Constructivists, by contrast, contend that scientific findings are substantially dependent on the culture and institutions that produce them. For constructivist, idealized notions of scientific production ignore the obstacles posed by cultural assumptions, religious teachings, interest group politics, disciplinary disputes, ideological beliefs, personal interactions, conflicts of self-interest, and the desire for fame, and the hanging sword of tenure. Realizing the impact of these factors, constructivists say, should lead one to displace the reification of scientific methodology with the realization that some, even most research findings result from social factors. To paraphrase Shelia Jasanoff, legal actors should abandon the notion that there are clear definitions of “good science” from which judges may separate inadequately “scientific” opinions from reliable expert testimony.

There is little evidence to suggest that either federal judges or members of Congress explicitly considered epistemology during the choice or creation of our admissibility regimes. The Daubert court addresses the issue only in a brief and haphazard manner, as I will argue below. Regardless, evidentiary regimes may still be

---

17 Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
19 AUGUSTE COMTE, THE POSITIVE PHILOSOPHY OF AUGUSTE COMTE (Thoemmes Continuum 2002).
20 PAUL FEYERABEND, AGAINST METHOD: OUTLINE OF AN ANARCHISTIC THEORY OF KNOWLEDGE (1975).
23 SHEILA JASANOFF, SCIENCE AT THE BAR: LAW, SCIENCE, AND TECHNOLOGY IN AMERICA (Harvard University Press 1995). Though this remains sound advice, it is unclear, exactly, how naïve judges are in practice about such claims. Though a simplistic reliance on tests such as those in the Daubert dicta might raise this critique, historical studies of judges and expert testimony suggest judges might not be so simplistic. See TAL GOLAN, LAWS OF MEN AND LAWS OF NATURE: THE HISTORY OF SCIENTIFIC EXPERT TESTIMONY IN ENGLAND AND AMERICA (Harvard University Press 2004).
24 Daubert, discussed infra Parts II.B., makes Popperian falsification as one of the foremost markers of “good science.” Though useful for rejecting tautologies, falsification is no longer seen as fundamental to the scientific method by modern philosophers of science. See, e.g., Caudill & Redding, supra note 18;
properly analyzed from an epistemological standpoint, since their tests and guidelines rely upon a particular vision of science even if their authors were not aware they were doing so.

Apart from epistemology, support for a particular admissibility regime also depends on whether one believes the courts have the institutional capacity to implement a particular epistemological vision. A realist might believe that the scientific method haltingly uncovers fundamental reality and yet argue that judges have little or no ability and willingness to assess this process. If judges—or the court system in general—lack the expertise or will to incorporate mainstream scientific analysis into admissibility decisions, a deferential evidentiary standard might be the only alternative.\(^{25}\) One of the best known statements of concern over an active gatekeeping role comes from Justice Rehnquist’s opinion (concurring in part and dissenting in part) in \textit{Daubert}:

> I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its ‘falsifiability,’ and I suspect some of them will be, too. I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony. But I do not think it imposes on them either the obligation or the authority to become amateur scientists in order to perform that role.

Rehnquist may sell judges short, particularly if the choice is framed not as judges or nothing, but instead as judges or juries. Judges are “repeat players” who will encounter similar claims of expertise over multiple occasions, and will almost certainly have more experience than a jury in evaluating the claims of expert witnesses.\(^{26}\) Moreover, as repeat players, judges have educational and developmental resources available to them that juries do not, as a judge who wishes to understand the PCR method...


for DNA typing may turn to the relevant article from the excellent *Reference Manual on Scientific Evidence*, or attend a workshop on the same.

*Daubert*, I believe, neither provides a clear picture of epistemology nor realistically weighs the burdens it imposes on judicial institutions.

B. *The Federal Standard: Daubert v. Merrell Dow Pharmaceuticals*

At present, American legal jurisdictions have generally adopted one of two admissibility regimes: either some version of the general acceptance test (the “*Frye*” regime), or some version of the federal *Daubert* decision. This leads to the tendency to label jurisdictions as “*Frye*” or “*Daubert*” courts, though this binary categorization admittedly conflates important differences. The *Frye* regime, taking its name from a 1923 federal criminal case in which the United States challenged the defendant’s proffer of a polygraph test as evidence of his innocence, simply asks whether the expert testimony at hand is “generally accepted” by the scientific community at hand. At that time, courts often judged the admissibility of expert testimony according to its utility in the marketplace; a technique, idea, or concept that was reliable enough to be consumed, used, or purchased was reliable enough for courts. Aside from the rather heroic assumptions such a test makes about consumers, this test had little relevance for forensic techniques, whose only “consumers” were law enforcement and the court system. Faced with a defendant who claimed a systolic blood pressure test could demonstrate his innocence, the appellate court ruled that this forerunner of the polygraph had not been generally accepted in the scientific community, and thus was not fit for admission. Though the *Frye* case was largely ignored for decades, it became a leading

28 David E. Bernstein & Jeffrey D. Jackson, *The Daubert Trilogy in the States*, 44 *Jurimetrics* J. 1 (2004). As he finds, a state might accept a *Daubert*-like admissibility test, but refuse to accept *Daubert’s* progeny, *Joiner* and *Kumho*. A state might also exempt certain kinds of evidence from an admissibility regime, or adopt different standards of admissibility for criminal and civil courts. Finally, a state might remain a “*Frye*” state but find its actual use of the general acceptance test altered by *Daubert’s* influence. Bernstein gives a thorough breakdown of these various categories.
29 *Frye* v. United States, 293 F. 1013, 3-4 (D.C. Cir. 1923).
31 As Faigman, Porter, and Saks note: “But though it might be practical and easily administered, the test of commercial value is a poor one. Its major weaknesses are perhaps more obvious today than they were a century or two ago. The market not only selects for validity, it selects also for entertainment, desire, wishful thinking, hope, sometimes even desperation. These are not without their value, but they are not good proxies for what courts are looking for in expert testimony. If the marketplace approves, as it does, of astrologers, sellers of phony cancer cures, and guides to new age vortexes, are those therefore good enough to provide guidance in a courtroom? The marketplace test is incapable of distinguishing astrophysics from astrology.” Faigman et al., *supra* note 16, at 1805.
candidate for the national admissibility regime in the 1970s, perhaps in reaction to the increasing debate over the use of scientific evidence in the courtroom.\textsuperscript{32}

The Frye regime transferred judicial deference from the marketplace to the scientific community. On the question of epistemology, Frye takes a simplistic realist position, putting faith in scientific communities to separate good science from bad. There is no recognition of the social and cultural factors that may affect scientific claims, perhaps in line with the greater influence positivism held in the academy at the time Frye was written. As for the question of institutional capacity, the regime largely absolves judges from gate-keeping, their only task to ascertain the extent of the relevant scientific community and the degree of consensus within that community. Frye is properly viewed as a “conservative” admissibility regime, both because it will take time for new expert testimony to be generally accepted by at least a plurality of the relevant expert community,\textsuperscript{33} and because once expert testimony has gathered a significant degree of consensus it will become hard to exclude.\textsuperscript{34}

Frye’s virtue is its simplicity and the modest demands it makes on judicial capacity. Its great vice, unsurprisingly, is that same simplicity. A general acceptance test tells us very little about the proper scope of the scientific community that constitutes that group whose acceptance is needed. All things being equal, a narrower group of scientists who are heavily invested in a technique or opinion will be quicker to support more controversial expert claims than a broader community. One can thus manipulate Frye by broadening the scope of the relevant community to the point where general acceptance is unlikely, or narrowing it to the point where the chances of agreement increase. The problem of scope worsens when addressing forensic testimony, where the expert community is likely limited to the practitioners themselves. On the one hand, fingerprint experts will be the most capable group of individuals for ascertaining what fingerprint techniques are generally accepted. On the other, the same experts are unlikely to entertain claims that their enterprise lacks scientific rigor or validity.\textsuperscript{35}

The stronger critique of Frye is that deference to expert agreement is no substitute for an actual analysis of evidentiary worth. The history of science is replete with expert communities who favored theories that scientists now agree lack validity.\textsuperscript{36} If one accepts that scientific results are at least in part socially constructed, that whether for cultural, religious, financial, or institutional reasons expert communities do not always simply work towards truth, then an admissibility regime of deference will inevitably admit unworthy expert testimony. Indeed, without some degree of independent gate-

\textsuperscript{32} Sanders, supra note 13, at 376-77; Saks, supra note 30, at 1073-77.

\textsuperscript{33} At least one scholar has argued that Frye is so restrictive that its use might more quickly lead to evidentiary reform in the statutory sphere. In other words, she would gamble that Frye would keep out so much useful and reliable expert testimony that Congress or state legislatures would be moved to institute a more liberal admissibility regime. Farrell, supra note 21, at 2214-2217. I am not so sanguine on this position.

\textsuperscript{34} Saks & Faigman, supra note 10, at 119.

\textsuperscript{35} Cole, supra note 8; Epstein, supra note 8; David Kaye, The Nonscience of Fingerprinting: United States v. Llera-Plaza, 21 QUINNIPAC L. REV. 1073 (2003).

\textsuperscript{36} For example, one might cite phrenology (the study of how head shape indicates character and intelligence), vile theories of racial difference, and arguments that one’s body measurements serve as indicators of criminal tendencies. That said, a defender of Frye might reasonably ask whether an active judicial gate-keeper, even within the crucible of an idealized adversarial process, would have been any more successful in exposing the problems with these theories.
keeping, the *Frye* test can become an absurdity. If a litigant tendered astrological evidence in support of a factual claim, would the community of astrologers be consulted for general acceptance? Or would the judge simply shift his analysis to astronomers or scientists in general to ensure the expertise is excluded? Either way, the example demonstrates the problem with equating expert agreement with validity, particularly for small or specialized expert communities. The problem exists even in more legitimate debates—if doctors and epidemiologists disagree about the proper methods of inferring causation, whose opinions should constitute the *Frye* analysis?

Apart from its theoretical problems, *Frye* faced a more immediate difficulty in federal courts following the 1975 promulgation of the Federal Rules of Evidence. *Frye*, viewed as a fairly narrow and generally conservative admissibility regime, seemingly conflicted with the Rules’ generally liberal attitude towards the admission of evidence. The problem exists even in more legitimate debates—if doctors and epidemiologists disagree about the proper methods of inferring causation, whose opinions should constitute the *Frye* analysis?

702 clearly required that expert testimony be relevant and helpful, and that the expert who tendered it be qualified. What was not clear, however, was whether the rule’s phrasing inferred a “reliability” requirement. On the one hand, its plain meaning did not present a reliability requirement. On the other hand, the rule’s use of the term “scientific knowledge” may reasonably be read to require some baseline of “scientific” quality in order to earn the name. The federal circuits reflected this split of opinion, as some rejected *Frye* as too restrictive in light of the Federal Rules, some read *Frye* into

---

37 Sanders, supra note 13, at 376-77.
38 FED. R. EVID. 702 (1975) (amended 2000). In 2000, Federal Rule 702 was updated, probably to bring it into accordance with the Daubert decision. The new text reads: “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.”
41 See, e.g., United States v. Baller, 519 F.2d 463 (4th Cir. 1975); United States v. Williams, 583 F.2d 1194 (2nd Cir. 1978). Notably, these early cases rejecting a more stringent admissibility test took place in the criminal realm.
Rule 702, aiming for a more restrictive standard, and at least one circuit, foreshadowing *Daubert*, developed its own standards for reliability.

In order to resolve this circuit conflict, the Supreme Court chose to hear *Daubert v. Merrill Dow Pharmaceuticals*, a Ninth Circuit appeal in which the Daubert family had sued Merrill (later purchased by Dow), alleging its anti-nausea drug Bendectin had caused serious birth defects in their unborn child. The *Daubert* case came at the tail end of a long series of Bendectin torts in both state and federal court, whose progress came alongside considerable legal, scientific, and media attention. While Bendectin plaintiffs had several early jury successes (Merrell made for a particularly unfavorable defendant, given its reticence to divulge its own research and its prior involvement with other unsafe drugs such as thalidomide and MER/29), ensuing epidemiological studies slowly but surely built a body of scholarship that found no demonstrable correlation between the use of Bendectin and birth defects. Both the trial court and the Ninth Circuit relied on the general acceptance test from *Frye* in their analysis; the Ninth Circuit also read into Rule 702 a peer review requirement. Finding the Dauberts’ expert testimony inadequate, the trial court ruled to exclude it, leaving the plaintiffs subject to summary judgment. The Ninth Circuit affirmed. The Dauberts appealed to the Supreme Court, arguing that the Ninth Circuit’s interpretation of *Frye* contradicted the liberal emphasis of the Federal Rules of Evidence, and that as statutory law the Rules superseded the common-law *Frye*.

The Supreme Court unanimously agreed with the Dauberts that Federal Rule 702 supplanted *Frye* as the standard for expert admissibility, arguing that the more rigid general acceptance test did not square with the “liberal thrust” of the Federal Rules. However, the Court also agreed with the *amicus* who supported Merrell Dow that Rule 702 imposed a reliability requirement. While the Ninth Circuit’s construal of reliability

---

42 See, e.g., Dang Vang v. Vang Xiong X. Toyed, 944 F. 2d 476 (9th Cir. 1991); United States v. Kozinski, 821 F.2d 1186 (6th Cir. 1987).
43 United States v. Downing, 753 F.2d 1224 (3d Cir. 1985).
44 *Daubert v. Merrell Dow Pharmaceuticals*, Inc., 951 F. 2d 1128 (9th Cir. 1991).
46 *Id.* Forty such epidemiological studies were done between 1963 and 1994. None of these studies found a significant correlation between the drug and the defects, though six found that Bendectin might be a teratogen but lacked sufficient evidence to draw a conclusion. While some of the early studies lacked adequate statistical power to be truly persuasive, the body of work became more sophisticated and convincing over time. Bendectin plaintiffs were increasingly forced to rely on animal studies and structural analyses to make their case for causation. Animal studies are dubious evidence of causation because of the difficulties of extrapolating animal results to humans, as well as because the animals in question often received atypical dosage of the substance in question (reactions to toxic chemicals are not always linear). Structural analyses compare the structure of known teratogens to those of the drug in question, but as minute differences in structure can lead to very different physiological reactions, their use in establishing causation is also limited.
47 *Daubert v. Merrell Dow Pharmaceuticals*, Inc., 951 F. 2d 1128 (9th Cir. 1991). The Dauberts’ expert witnesses had either reanalyzed previously unsupportive epidemiological data or relied upon animal studies and case reports as the basis for their testimony. The trial judge ruled that their methods were not generally accepted by epidemiologists for assessing causation (though they were acceptable for risk assessment by a body such as the EPA), and that their opinions had not been peer-reviewed.
was too narrow, Blackmun agreed that FRE 702 bestowed a “gate-keeping” responsibility on the trial judge to ensure all expert testimony was reliable, relevant, helpful, and “fit” the issue at hand.50

The most important task for federal judges, uneasy with their new responsibilities,51 was parsing Blackmun’s definition of “reliability.” Rightly contending that methodological quality defied a straightforward definition, Blackmun said that the judicial assessment of reliability should be flexible, and refrained from creating tests or rules that would constrain how reliability should be judged. However, perhaps recognizing that many federal judges would need some sort of guidance, he suggested four criteria as dicta to guide lower courts in their assessment.52 These “Suggestions” were as follows:

1) Can the opinion be falsified?
2) Has the opinion appeared in a peer-reviewed journal?
3) Does the opinion present an error rate?
4) Is the opinion generally accepted among the scientific community?

A brief examination of these factors shows the influence of the philosopher Karl Popper,53 the importance of peer review despite its rejection as a formal requirement,54 and the return of Frye, though in a supporting role rather than a leading one.55 Despite

---

50 The last criterion is somewhat ambiguous, but has generally been held to mean that the expert testimony must not only be relevant but properly applied to the facts at hand. In Blackmun’s words: “The study of the phases of the moon, for example, may provide valid scientific knowledge about whether a certain night was dark, and if darkness is a fact in issue, the knowledge will assist the trier of fact. However, evidence that the moon was full on a certain night will not assist the trier of fact in determining whether an individual was unlikely to have behaved irrationally on that night.” Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 591 (1993). Quoting Judge Becker from In re Paoli, 35 F.3d 717 (3d Cir. 1994), Sanders helpfully points to Judge Becker’s opinion from In re Paoli: “[A]nimal studies may be methodologically acceptable to show that chemical X increases the risk of cancer in animals, but they may not be methodologically acceptable to show that chemical X increases the risk of cancer in humans.” Joseph Sanders, supra note 13, at 378 n.46.
51 Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311, 1316 (9th Cir. 1994).
52 When one compares the text of the Solicitor General’s brief with the Court’s opinion, the influence of the former is hard to miss. Brief for the United States as Amicus Curiae Supporting Respondent, Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1992) (No. 92-102).
53 For a summary of what some philosophers of science see as Daubert’s over reliance on falsification, see Edmond & Mercer, supra note 24.
55 The Daubert paradigm was clarified by two further Supreme Court decisions: General Electric Co. v. Joiner, 522 U.S. 136 (1997), which explained the appropriate standard of review for Daubert hearings, and Kumho Tire Co. v. Patrick Carmichael, 526 U.S. 137 (1999), which answered whether Daubert was
the admonition that *Daubert* analyses should remain flexible, some judges understandably rely on these dicta as a four-part test to be applied in the face of unfamiliar expert testimony.\(^{56}\)

**C. Daubert and the Theoretical Framework: Realist, Constructivist, or Neither?**

How does the *Daubert* regime treat the foundational concerns of epistemology and judicial capacity? On epistemology, *Daubert* is simply ambiguous. On one hand, Blackmun’s four indicators of “good science” expressly ratify the realist position that the proper use of scientific methodology is the gateway to reliable evidence and accurate decisions. On the other hand, several phrases in the opinion suggest a more constructivist viewpoint.\(^{57}\) For example, the Court noted imperfections in the very tests it would later promulgate, citing Jasanoff’s work on the failures of peer-review,\(^{58}\) and praised cross-examination as a tool for exposing inconsistencies that scientists—presumably bound by a particular paradigm or their own institutional and scientific interests—might not catch. The adoption of judicial gate-keeping in *Daubert* is itself a nod towards the constructivist position, since it proceeds on the expectation that the litigants’ experts cannot readily separate their scientific claims from their own financial interests.\(^{59}\)

*Daubert*’s epistemology thus appears a mix of realism and constructivism, though more as the result of “muddling through” than a deliberate choice by the majority.\(^{60}\) Certainly, a trial judge who wished to find support for a realist or a constructivist epistemology within *Daubert* could do so. This ambiguity has been criticized as harmfully incoherent,\(^{61}\) as well as accepted (if not praised) on the grounds that judges intended to apply to only “scientific” evidence, or to all expert testimony. Together, the three cases are sometimes referred to as the “Daubert Trio.” Margaret A. Berger, *The Supreme Court’s Trilogy on the Admissibility of Expert Testimony, in Reference Manual on Scientific Evidence* ed., Second Federal Judicial Center (2000).

\(^{56}\) Dixon & Gill, *supra* note 7 at 282. Less than half of *Daubert* decisions go through exactly such a format, as Dixon and Gill found that many judges still relied on more traditional legal assessments of admissibility, such as prejudice or relevance, or ignored the requirement altogether. *See also* Jennifer L. Groscup et. al., *Expert Testimony in the Courts: The Influence of Daubert, Joiner, and Kumho Decisions, Part 3: The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Courts*, 8 PSYCHOL. PUB. POLY. & L. 339 (2002).

\(^{57}\) Caudill & Redding, *supra* note 18, at 736-742; Farrell, *supra* note 21, at 2198-2208.


\(^{59}\) Later refinements of *Daubert* increased the formal recognition of constructivist factors. After the Court remanded the Daubert case for reconsideration under the new standard, the trial judge again excluded the plaintiff’s experts under *Daubert*. On appeal, the Ninth Circuit upheld the exclusion, and added to the *Daubert* dicta a further consideration: expert opinions that arose naturally from a scientist’s work should be viewed as more reliable than those which arose during the trial process. The constructivist insight here is straightforward—while experts should be equally capable of following the scientific method and doing good research when their salary depends on the results turning out a certain way, we are rightly skeptical of claims made by persons in that situation.


\(^{61}\) Farrell, *supra* note 21, at 2213-2217.
have neither the time nor the ability to resolve debates about the nature of science when the field itself lacks consensus.\textsuperscript{62}

On the question of judicial capacity, \textit{Daubert} may sometimes require more of federal judges than federal judges have to give, raising problems the \textit{Frye} regime sidestepped. \textit{Daubert} tells judges, to paraphrase Michael Saks, to “figure it out for yourselves.”\textsuperscript{63} Beyond Rehnquist’s concern that judges will misunderstand concepts such as falsification, Learned Hand’s dictum remains in full effect. Discerning potential methodological shortfalls in the aforementioned Bendectin litigation would require that the judge have a working knowledge of relative risk, in-vitro studies, epidemiology, structural analysis, statistical significance, the fallacies of temporal association, and the risks of extrapolating from animal studies to human beings. And that covers merely one portion of the substantive expertise in which a generalist trial judge, who likely has had no formal scientific training, must assess the reliability of expertise. Needless to say, one does not need to question the intelligence of federal judges to wonder whether they are up to the task.\textsuperscript{64} Questions of institutional capacity may be mitigated by repeated exposure to similar \textit{Daubert} motions. Federal judges also have access to training seminars, resources such as the excellent \textit{Reference Manual on Scientific Evidence},\textsuperscript{65} though it remains an open question whether they adequately take advantage of such opportunities.\textsuperscript{66} My personal view of the record is mixed: some \textit{Daubert} opinions are humbling to the reader, while others should be humbling to their authors.

At present, my point is simply that as read, \textit{Daubert}’s text does not choose sides in either the ongoing debate between realism and constructivism or in the argument as to whether admissibility should be a liberal or conservative standard. In the following section, I will explain the likely results of this ambiguity, namely the enhanced autonomy judges gain when applying a precedent lacking both a rule-like structure and a clear message.

III. THE EMPTY VESSEL: \textit{DAUBERT} AND THE LOCAL CONSTRUCTION OF RELIABILITY

A. Law, Legal Realism, and Judicial Decision-Making

Since the early 20\textsuperscript{th} century, political and legal scholars have hotly debated the degree to which the law actually constraints, guides, or shapes legal decision-making. As with arguments regarding epistemology, the range of debate is book-ended by a pair of strawman arguments. On one side, judges mechanistically interpret statutes and apply legal rules, finding the law without making it. On the other, judges are no more than legislators in black robes, free-wheeling creators of policy who use the law as a cloak to blur their naked use of power. Again moving from the poles of the continuum to the

\begin{itemize}
  \item \textsuperscript{62} Caudill & Redding, \textit{supra} note 18, 762-767.
  \item \textsuperscript{63} Sanders, \textit{supra} note 13, at 379.
  \item \textsuperscript{64} Atikian, \textit{supra} note 25; Paul S. Milich, \textit{Scientific and Technological Evidence: Controversial Science in the Courtroom: Daubert and the Law’s Hubris}, 43 \textit{EMORY} L.J. (1994).
  \item \textsuperscript{65} \textit{REFERENCE MANUAL ON SCIENTIFIC EVIDENCE}, \textit{supra} note 27.
  \item \textsuperscript{66} A survey of Texas judges, for example, found that seventy percent of the responding judges took “no continuing education or practical business experience in the use and analysis of the reliability of scientific methodology.” Montz, \textit{supra} note 25.
\end{itemize}
more reasonable middle, one finds two basic schools of thought. Some distance from the latter extreme, one finds the proponents of the “attitudinal model,” inheritors of the legal realist school who argue policy preferences and ideology are the single most important factor in judicial decision-making at the Supreme Court level. Their theory is a simple one: laws, constitutional clauses, and precedent are sufficiently indeterminate to allow judges to decide cases as they wish. This premise of legal realism, it should be noted, rests the lack of meaningful appellate oversight. The attitudinal model works best in situations where institutional constraints are minimized, such as at the Supreme Court, where justices with life tenure and no superiors have little to fear from making any particular decision.

Though the notion that judicial decision-making has a strong ideological component is a consensus position among empirical legal scholars (and perhaps by any serious observer of courts since there have been courts), critics of attitudinalists argue the realists have overreached, inaccurately reducing judicial decision-making to a single dimension. This perceived tendency of attitudinalists to overreach has led other scholars to examine alternative models of decision-making, such as examining how judges make strategic decisions to overcome collective action problems, or how court institutions exert an independent effect on decision-making. One promising approach which focuses on precedent as a type of institution views landmark rulings not as rules that constrain behavior, but rather as “legal regimes” that constitute the frameworks and boundaries of decision-making. The authors of this approach have convincingly argued that the Supreme Court voted differently on core speech cases following the creation of the “two-track” content-neutrality and content-based regime in Police Dept. of Chicago v. Mosley and Grayned v. Rockford. In other words, while the content-based versus content neutral regime does not dictate the result of particular cases or remove the ideological component of such decisions, it has changed the manner in which the Justices

---

67 JEFFREY A. SEGAL & HAROLD J. SPAETH, THE SUPREME COURT AND THE ATTITUDINAL MODEL (Cambridge University Press 1993). The book can be summed thus: “Rehnquist votes the way he does because he is extremely conservative; Marshall voted the way he did because he is extremely liberal.” Id, at 65. It should be noted that the initial claims of attitudinalists—that political ideology overwhelmingly explains judicial behavior—have been dialed down in more recent years. Ideology is likely the most influential factor in Supreme Court decisions—particularly in constitutional and civil liberties case—but by itself only explains roughly one-fifth to one-third of such decisions. Wilson, Sean, The Attitudinal Model, Political Science, Ecological Fallacy and Exaggeration (August 3, 2006). Available at SSRN: http://ssrn.com/abstract=922183.


69 SEGAL & SPAETH, supra note 67.


73 Police Dept. of Chicago v. Mosley, 408 U.S. 92 (1972)

approach and consider core speech cases, in turn affecting the general pattern of outcomes. In a similar (albeit qualitative) vein, Howard Gillman persuasively contends that legal ideology regarding the proper use of state police powers (here seen as distinct from the standard liberal-conservative continuum along which attitudinalists situate their models) heavily influenced amendment jurisprudence between the Civil War and the New Deal. Either a landmark precedent or a persuasive legal ideology, therefore, may temper more the standard liberal and conservative positions that enter the courts through the appointment process.

How likely is Daubert to affect judicial decision-making, given what we know about judicial decision-making? From a greatly truncated description of a vast literature, let me draw the following three observations. First, institutional context matters. An appellate court judge, for example, will feel far more pressure to adhere to precedent than a Supreme Court justice. Second, important constitutive precedents such as Mosley, or to borrow from Kritzer and Richards, “legal regimes,” may have greater influence on lower court decisions when they are formalistic. If precedents and prior interpretation do constitute the grounds for debate and decree the mode of analysis, it seems reasonable to assume that the grounds for debate—and thus the degree of autonomy for judges—will be broader under a balancing or a totality of the evidence analysis than under a more formal method, such as the three-part test found in Brandenburg v. Ohio.

Third, textual or doctrinal ambiguity may be offset by shared visions of legal ideology or legal goals, goals distinct from (if often overlapping with) liberal or conservative policy agendas. In other words, law does not consist simply of doctrine and tests, but also a shared “state of mind” among legal practitioners. To draw an example from constitutional law, the chance of a racially discriminatory law surviving strict scrutiny is low not only because the difficulty in crafting a law that is sufficiently narrowly tailored to address a compelling interest, but also because of the shared understanding of post-Warren Court judges that such laws are generally illegitimate. Likewise, interpretations of the amendment have depended greatly on competing understandings of whether the Court should demarcate and enforce the boundaries of federalism, or whether those boundaries should be set by voters. Once rule-based federalism or political federalism gained currency among a majority of the Court, that

75 Richards and Kritzer have successfully replicated this approach for search and seizure Supreme Court decisions centered on Mapp v. Ohio, as well as establishment clauses Supreme Court decisions centered on Lemon v. Kurtzman. Herbert M. Kritzer & Mark J. Richards, The Influence of Law in the Supreme Court's Search-and-Seizure Jurisprudence, ” American Politics Research 33 AM. POL. RES. 33 (2005); Herbert M. Kritzer & Mark J. Richards, Jurisprudential Regimes and Supreme Court Decision Making: The Lemon Regime and Establishment Clause Cases, 37 LAW & SOC'Y. REV. 827 (2003).


77 Richards & Kritzer, supra note 72.

78 See Guthrie, supra note 11; ADRIAN VERMEULE, JUDGING UNDER UNCERTAINTY: AN INSTITUTIONAL THEORY OF LEGAL INTERPRETATION (Harvard University Press 2006).


philosophy animated future doctrine, further clarifying what the Court expects from lower courts. A legal regime that creates such a shared understanding should constraint or guide judicial decision-making more than a precedent whose meaning is contested, or worse, unclear, regardless of the rule, test, or analysis involved.

B. The Institutional Context of Daubert Decisions

How do these considerations apply to admissibility decisions under the Daubert regime? As far as institutional context goes, federal judges making admissibility decisions operate largely free from the fear of reversal. If Daubert decisions do reach an appellate court, the legal standard for appellate review is whether the judge’s decision was an “abuse of discretion”, a deferential rule. The appellate judge clearly has some wiggle room in deciding what constitutes such an abuse, but the Supreme Court has made it clear in Joiner that it expects significant deference from its appellate courts. Moreover, admissibility decisions may lack the salience that would lead to active policing of Daubert decisions by busy circuit judges, especially given that appellate judges may correctly determine that trial judges are “closer to the facts” than they are. One recent study of Daubert decisions in federal torts supports this theory, finding no evidence that judges act strategically to avoid being overturned by appellate courts.

A quick tabulation of the affirmance rate of federal trial-level evidentiary decisions also supports this hypothesis. Between 2002 and 2005, for example, the affirmance rate of appellate courts for Daubert decisions at the trial sat at just under ninety percent. Of course, this high affirmance rate does not count the numerous decisions that are never appealed. While judges may not be aware of this exact rate, they are surely aware of the magnitude of its effect, making them secure that their admissibility determination is unlikely to be challenged. In short, the institutional context for admissibility decisions is one of considerable autonomy. Judicial autonomy, in turn, permits wide variation among judges in interpreting and implementing Daubert in different contexts.

C. Daubert’s Fractured Stance on Admissibility

1. Admissibility in the Daubert Opinion

How does Daubert fare in providing doctrinal guidance for trial judges? Here again, the nature of the Daubert regime increases the probability of fragmentation, for as doctrine, Daubert is severely underdetermined. The decision not only eschews formalism; it explicitly denies that its gatekeeping responsibility can be formulated in any

83 Id.
84 Buchman, supra note 7, at 682-88.
85 Peter J. Neufeld, Temporal Trends (?) in Appellate Review of Daubert Decisions (2005), http://www.daubertontheweb.com/2005/02/temporal-trends-in-appellate-review-of.html. There is a slight chance—as Neufeld concedes—that a few appellate decisions may have been missed because the opinion lacked the term “Daubert.” However, even if these omissions are slanted in one direction and affect the win rate, Neufeld argues that as probable trite affirmations, their inclusion would increase, not decrease, the win rate.
sort of rule-like language. Such avoidance seems reasonable from a philosophic standpoint, as asking a legal test to duplicate the “scientific method” when there is little agreement on exactly what that comprises is a doomed enterprise. Nevertheless, as empty doctrine, *Daubert* essentially makes each judge the master of his own fiefdom, absent on-point precedent from the relevant circuit. The lack of doctrinal clarity should also decrease the chance of reversal (again absent on-point precedent), as finding an abuse of discretion may be more difficult to justify when the standards are hopeless vague.

*Daubert* observers might contend that Blackmun’s dicta—the “considerations” of testing, peer-review, error rate, and general acceptance—serve as a rough formal test that mitigates these effects. A harried judge might simply ask whether the expert’s testimony is testable, is peer-reviewed, presents an error rate, is generally accepted, and be done. Two objections prevent this claim from gaining traction. First, while use of the *Daubert* dicta remains more common than the use of other indicia of arguments about reliability, they appear in less than half of all such admissibility decisions.86

Frequency of use aside, even if one treats the dicta as doctrine, the *Daubert* opinion provides little aid on their interpretation. Judges may not even understand concepts such as falsifiability,87 especially if comprehension requires not only adequately defining the term but understanding the philosophic debates regarding its application.88 The difference between one reasonable interpretation and another might mean the difference in admitting or excluding the evidence before the court. For example, does testing and falsification require that the expert’s opinion have been actually tested? That it is capable of being testing? Or does a history of prior admission in a federal court serve as “adversarial testing” that satisfies the criterion?89 Similarly, does peer review require publication within “an unbiased and financially disinterested community of practitioners”90? Or may peer review literally require only that expert opinions be reviewed by one’s physical peer, as is the case for fingerprint technicians?91

Moving to error rates, does the error rate prong require a formal probability, as with DNA cases? Or can it be satisfied by the expert’s (laughable) claim that the error rate of a particular technique is zero?92 Finally, must general acceptance analysis extend beyond the agreement of a small expert community whose financial and employment interests are bound up in a technique’s acceptance? Or is general acceptance intended to examine what technique that particular community supports?93 Judges may offer

87 Montz, *supra* note 25. Montz surveyed a number of state judges, finding their understanding of falsification, for example, was limited at best. It is possible that federal judges have a better understanding of such terms than their state counterparts.
91 See, e.g., Havvard, 117 F. Supp. at 854.
93 Cole, *supra* note 8, at 1216, 1221, 1245; Epstein, *supra* note 8, at 646.
significantly different interpretations of these dicta, with little guidance from the *Daubert* decision or the Supreme Court as to which interpretations are more appropriate. Such differences make it unlikely that use of the dicta as a rough and ready doctrinal test could elucidate an analysis the Court has chosen to make opaque.

*Daubert*’s vacuous doctrine could still be limited by signals from the Supreme Court that it desired a general position on the admission of experts, whether liberal or restrictive. Blackmun’s recently released papers suggest the Court did intend to use *Daubert* to clarify the proper stance on admissibility. Despite such intent, his opinion raises more questions than it answers. *Daubert* supports both a liberal and a restrictive judicial stance on admissibility, depending on what parts of the opinion one cites. In his attempts to mollify both the vociferous critics of “junk science,” as well as those who worried that a reliability standard would “sanction a stifling and repressive scientific orthodoxy,” Blackmun equivocated on whether the *Daubert* standard was a liberal or a conservative one.

Support for a liberal standard, for example, can be found in Blackmun’s assurance that removing the Ninth Circuit’s general acceptance and peer-review requirements as the formal standard of admissibility would not lead to “a ‘free-for-all’ in which befuddled juries are confounded by absurd and irrational pseudoscientific assertions.” Blackmun instead counseled that the traditional tools of the adversarial system, such as cross-examination and jury instructions would suffice to supplement *Daubert* analysis. Blackmun’s support for the jury system and traditional legal safeguards against unreliable evidence, as well as his position that a liberal standard of admissibility was most consistent with the spirit of the Federal Rules of Evidence, convinced some legal scholars that *Daubert* ultimately took a “liberal” stance on admissibility. Similarly, in executing their gatekeeping responsibilities in the years immediately following *Daubert*, several district courts specifically noted the “liberal” nature of the Federal Rules and the *Daubert* decision in making their decision.

However, Blackmun’s opinion also supports a more conservative interpretation of the gate-keeping responsibility. Unlike scientific arguments, legal disputes must be resolved “finally and quickly,” Blackmun says, and hypotheses that are “probably wrong” will “be of little use” to courts. Moreover, Blackmun accepted Merrell Dow’s

---

94 Email from Michael Saks, to Assistant Professor of Government Robert Robinson, University of Alabama-Birmingham (Sun, Jul 1, 2007 at 9:25 PM) (on file with Professor of Law & Psychology Michael Saks, and Faculty Fellow, Center for the Study of Law, Science, & Technology)
95 HUBER, supra note 18.
97 Id. at 595.
98 Id. at 596.
101 Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 597 (1993). While Blackmun is correct that legal and scientific practice differ in their methods for ascertaining truth, he greatly over-idealizes scientific practice in this statement. Funding, tenure, and scholarly competition, for example, put rather discrete time limits on scientific exploration in the short-term, just as path dependency, cultural paradigms,
position that the word “science” in Rule 702 obligates judges to apply some reliability analysis, one presumably based on traditional realist criteria.\footnote{Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 589-95 (1993).} Given this apparent elevation of judges to gatekeepers of reliability—a view itself strengthened by Rehnquist’s explicit concern over the same—other legal scholars viewed \textit{Daubert} as a establishing a more conservative stance than had been true before it. Such authors either celebrated that federal courts would now take a harder look at “junk” testimony plaguing civil and criminal cases\footnote{David E. Bernstein, \textit{The Admissibility of Scientific Evidence After Daubert v. Merrell Dow Pharmaceuticals}, 15 \textit{CARDOZO L. REV.} 2139 (1994); Lester Brickman, \textit{On the Relevance of the Admissibility of Scientific Evidence: Tort System Outcomes are Principally Determined by Lawyers' Rates of Return}, 15 \textit{CARDOZO L. REV.} 1755 (1994); Randolph Jonakait, \textit{The Meaning of Daubert and What That Means for Forensic Science}, 15 \textit{CARDOZO L. REV.} 2103 (1994).} or worried that this stance would raise the costs of bringing a case to trial, aiding corporate defendants more than anything else.\footnote{Roisman, \textit{supra} note 99.} Support for a more restrictive position on admissibility can also be found in the second and third cases of the “\textit{Daubert trio},” \textit{G.E. v. Joiner} and \textit{Kumho Tire v. Carmichael}. While \textit{Joiner} and \textit{Kumho} are better known for their clarifications of \textit{Daubert} standards than their factual situations, the Court’s own application of its standards to the case before them better fit a conservative stance than a liberal one.\footnote{General Electric Co. v. Joiner, 522 U.S. 136, (1997); Kumho Tire Co. v. Patrick Carmichael, 526 U.S. 137, (1999); Berger, \textit{supra} note 54.}

The split regarding \textit{Daubert}’s position on admissibility has, after almost fifteen years, given way to more nuanced arguments about \textit{Daubert}’s admissibility standards being liberal or conservative in particular contexts. For example, one developing consensus views federal courts as taking a relatively liberal position in criminal cases (particularly where the state’s evidence is concerned) while being much more restrictive in civil cases.\footnote{Rehnquist’s opinion in \textit{Joiner} required that each piece of the plaintiff’s expert testimony be reliable in its own right, rather than reliable as a whole. \textit{Joiner}, 522 U.S. at 141-48. Breyer’s opinion in \textit{Kumho} finds the plaintiff’s expert to be qualified, and his method of visually inspecting tires to look for defects a reasonable one, but rejects his argument that because no other causes for the blowout can be found, a manufacturing defect must be the cause. \textit{Kumho}, 526 U.S. at 153-59. These two types of arguments—requiring that each piece of evidence be adequate evidence of causation on its own, and requiring that an expert “rule in” a factor as a cause by substantive evidence rather than simply “ruling out” other potential causes—have frequently been used to exclude plaintiff expert testimony under \textit{Daubert}. See Grimes v. Hoffmann-LaRoche, Inc., 907 F. Supp. 33 (D.N.H. 1995) (for an example of both such arguments).} Another possibility is that federal courts “grandfathered” previously admitted expert testimony by providing it with a liberal standard, while requiring genuinely new expert testimony to meet a more restrictive standard.\footnote{See, e.g., Berger, \textit{supra} note 7; Dixon & Gill, \textit{supra} note 7; Giannelli, \textit{supra} note 8; Neufeld, \textit{supra} note 8.} In any case, the possibility of \textit{Daubert} creating a global shared understanding regarding admissibility seems remote. As with doctrinal clarity and institutional context, the Supreme Court’s own ambiguous position on admissibility, as well as division among influential commentators, contributes to fragmentation and the local construction of reliability.

2. Contrasting Admissibility Standards in *United States v. Crisp*

As an example of the degree to which divergent positions on expert admissibility may be taken from *Daubert*, I will discuss *United States v. Crisp*, a Fourth Circuit appeal challenging the admission of the government’s fingerprint and handwriting comparison experts. \(^{109}\) Crisp had been convicted of bank robbery, largely on the strength of accomplice testimony, latent prints drawn from palm prints, and handwriting comparisons of a note Crisp allegedly wrote to another accomplice. The fingerprint and handwriting expertise was central to the verdict. As the majority concedes, “the credibility of both Mitchell and Torain [the co-conspirators] was substantially impeached,” and the government itself conceded that the expert testimony was necessary to prove Crisp’s guilt beyond a reasonable doubt.\(^{110}\)

Chief Judge Wilkins joined Judge King in affirming the district court’s decision to admit the government’s experts, while Judge Michael dissented. More striking than the split decision, however, was the manner in which each opinion envisioned the *Daubert* gate-keeping responsibility. King began by citing some recent precedent in which fingerprint expertise had been admitted under *Daubert*, as well as placing significant importance on the fact that fingerprint evidence had been admissible for decades. Moving to the *Daubert* dicta, King applied these criteria to the state’s evidence in an extremely deferential fashion. Referencing a similar opinion in *United States v. Havvard*,\(^ {111}\) the majority argued that despite its lack of empirical testing, fingerprint evidence and handwriting evidence had been sufficiently “tested” in the adversarial system. In other words, if there were methodological problems with such evidence, or if their premises were false, surely several decades worth of cases would have caught them.

King then glided over the peer review requirement, only citing *Havvard*’s statement that an individual technician’s work was “reviewed” by one of his peers.\(^ {112}\) When addressing the error rate requirement, King’s opinion simply accepted at face value the risible claim by the government that fingerprinting has an error rate of “zero.” Given that even DNA identification presents non-zero error rates, no intellectually honest judge should accept such a claim.\(^ {113}\) Finally, addressing the general acceptance prong, the majority noted that:

> While the principles underlying fingerprint identification have not attained the status of scientific law, they nonetheless bear the imprimatur of a

---

109 United States v. Crisp, 324 F.3d 261 (4th Cir. 2003).
110 *Id.* at 264 n.1.
111 United States v. Havvard, 260 F. 3d 597, 600-01 (7th Cir. 2001).
112 *Id.*
113 Crisp, 324 F.3d at 269. One response to this claim is that while the technique itself produces no errors, individual examiners may, of course, err. However, since the abandonment of the “points” system and the movement towards the more subjective ACE-V technique pioneered by Ashbaugh, the examiner essentially is the technique. DAVID ASHBAUGH, QUANTITATIVE-QUALITATIVE FRICTION RIDGE ANALYSIS: AN INTRODUCTION TO BASIC AND ADVANCED RIDGEOLOGY (CRC Press 1999). Much like a harbor pilot, one cannot readily separate the examiner from the technique. The same is true for handwriting expertise, where even the majority opinion in *Crisp* implicitly concedes that there is no uniform method or set of standards.
strong general acceptance, not only in the expert community, but in the
courts as well.\footnote{Crisp, 324 F.3d at 268.}

Expert communities whose interests and incentives depend entirely on the acceptance of
a particular technique do not present a good measure of its reliability or validity. Worse,
the position that general acceptance can rest on acceptance by courts in effect
“grandfathers” certain techniques, exempting them from actual gatekeeping.\footnote{Cole, supra note 8.}

The most arresting fact about the majority opinion, however, is its assumption that
Crisp, rather than the government, bears the burden of proof. Confronted with the
accusation that the foundations and methods of fingerprint comparison lack adequate
empirical testing, Judge King responded that “Crisp cites no studies demonstrating the
unreliability of fingerprinting analysis.”\footnote{Crisp, 324 F.3d at 267.} Wrapping up its discussion of the Daubert
dicta, King wrote that:

While Crisp may be correct that further research, more searching scholarly
review, and the development of even more consistent professional
standards is desirable, he has offered us no reason to reject outright a form
of evidence that has so ably withstood the test of time.\footnote{Id. at 269.}

King then framed Daubert as a liberal standard, citing Blackmun’s dictum that “vigorous
cross-examination, presentation of contrary evidence, and careful instruction on the
burden of proof are the traditional and appropriate means of attacking shaky but

Why is this opinion so deferential? King showed his hand when he accused that
“Crisp today advocates the wholesale exclusion of a long-accepted form of expert
evidence… such a drastic step is not required of us under Daubert, however, and we
decline to take it.” A few lines later, he similarly wrote that “to postpone present in-court
utilization of this bedrock forensic identifier pending such research would be to make the
best the enemy of the good.”\footnote{Crisp, 324 F.3d at 270.} King clearly feared the institutional consequences of
excluding forensic evidence regularly utilized by the state; the degree to which this fear
influenced his decision is, of course, unknown.

By contrast, the dissent put teeth into the Daubert dicta and found the
governments’ expertise wanting. Michaels argued that viewing the test prong as
adversarial testing, or the view of general acceptance as acceptance in court misconstrues
the judiciary’s gatekeeping responsibilities: “Our adversarial system has much to
commend it, but it is not a general substitute for the specific Daubert inquiry.”\footnote{Id. at 273.} If, on

\footnote{Id. at 273.} Michaels notes that at its inception, fingerprint testimony was admitted largely without
challenge, allowing it to build a reputation of reliability that no lawyer—particularly no lawyer representing
an indigent like Crisp—would directly challenge (or even know to challenge). For a critical history of the
the other hand, testing requires actual empirical testing, peer review requires publication in peer-reviewed journals, and general acceptance requires acceptance beyond an expert community whose livelihood would be threatened by exclusion, Michaels said, the prosecution’s experts fail on all counts. The foundations of these two techniques—such as the probability that a latent print might match more than one full set of exemplars, or whether one’s handwriting is fixed, unique, and capable of individuation—have not been tested, nor have their propositions been put forward for review in scientific (rather than trade) journals. Some semblance of an error rate might be found in the discussion of proficiency tests, Michaels noted, but the particular tests cited either lacked external validity (i.e. they were too easy), or only marginally established expert proficiency.  

Concluding that the government had failed to satisfy *Daubert*, Michaels voted to exclude both sets of testimony. “The Government has had ten years to comply with *Daubert*. It should not be given a pass in this case.”

It would be tempting to conclude that the dissent has applied the “true” meaning of *Daubert* and the majority has misread it; tempting, but mistaken. The majority is certainly correct on one count: its approach to *Daubert* squares with the manner in which most federal courts have treated most attacks on the government’s experts in criminal cases. How could the majority federal decisions overwhelmingly advance the “wrong” interpretation of *Daubert* and suffer no consequences? While I would argue the majority’s interpretation of terms such as “testing” and “general acceptance” is laughably weak, King’s insistence that the *Daubert* court did not intend a radical restructuring of expert testimony in criminal cases may be correct. In any case, the point is not that one side or another applied the “correct” interpretation of *Daubert*, but instead that *Daubert* has no correct interpretation. Rather than create a clear and consistent stance on the standard by which trial courts should admit expert testimony, the *Daubert* regime allows individual courts or circuits to develop localized understandings of admissibility.

**D. Daubert’s Fractured Stance on Epistemology**

A final factor that might unify *Daubert*’s admissibility regime is the Supreme Court’s stance on epistemology, or in other words, what constitutes good science. As with its doctrine and position on admissibility, however, *Daubert*’s stance on epistemology can justify multiple interpretations. In one scholar’s terms, *Daubert* encourages a “pragmatic constructivism” about epistemological issues, sidestepping direct consideration of epistemology and inviting lower courts to do the same. Even were one to read into *Daubert* a particular philosophy of science, however, I argue here that epistemology is likely to fade before other, more salient judicial goals. Unlike Gillman’s depiction of early 20th century federal judges, who were sincerely convinced that the well-being of the Republic depended on constraining the use of the police power.

---

122 Crisp, 324 F.3d at 272.
123 See Neufeld, *supra* note 8; Risinger, *supra* note 8.
to the public good, it is hard to see judges spending a great deal of capital or placing particular emphasis on championing a particular epistemology.

As argued in Section II, *Daubert* does not rest on a single point along the realist-constructivist continuum. The language of the *Daubert* opinion gives judges considerable freedom to employ either realist or constructivist factors as they see fit. Why would the Court endorse such ambiguity, given that its apparent goal was to resolve circuit conflict over the appropriate standard? First, frankly put, engaging the philosophy of science is a difficult task not only because it is complex, but because there is no consensus position on what constitutes good science. When constructivists criticize courts for naively relying on realist indicia of reliability, they contend there is no toolbox of methods that will readily allow judges to demarcate the wheat from the chaff. In their view, deference to realist criteria not only creates an unrealistic and unattainable vision of scientific certainty, but also leads courts to miss their own part in its social construction.

In due turn, however, such critics overlook that the conclusions one might draw from their own work is equally varied and conflicting. How would a generalist judge, unlikely to have any training in epistemology or the sociology of science, discriminate between the views of Bruno Latour, Peter Feyerbrand, Peter Galison, and Thomas Kuhn in appropriating the “right” lessons from the sociology of science? Asking judges to take a “constructivist” stance towards science is itself a request with little internal consistency. If anything, such constructivist arguments are likely to serve as tools for pragmatic judges wishing to attack specific realist claims in particular situations. The judicial capacity for evaluating and discriminating higher-order epistemological arguments is arguably no greater than the judicial capacity for evaluating high-level scientific disputes. It is difficult, in my view, to blame the Supreme Court for not jumping headfirst into such murky debates.

Second, even were a judge to take a relatively principled “realist” or “constructivist” stance on scientific epistemology from *Daubert*, choosing their favorite philosophers of science to emulate, such a stance might not correlate with any expected position on the admissibility of expert testimony. Constructivist and realist positions may justify *either* a liberal *or* a conservative approach to the admission of expert testimony. Ultimately, the decision to admit challenged expert testimony is a decision about which

125 Gillman, supra note 76.
127 See Caudill & Redding, supra note 18.
129 Edmond and Mercer, who surveyed judges’ use of various important philosophers of science found, unsurprisingly, that such use demonstrated little understanding of that field’s development over the last few decades. See Edmond & Mercer, supra note 25. Specifically, the authors track the heavy influence of Karl Popper, whose work is largely responsible for the falsification prong of the *Daubert* “test,” but whose importance in the actual philosophy of science has faded. By contrast, recent leading lights of the field were ignored. The authors conclude that the sociology of science has at best failed to permeate the law, and at worst has been misappropriated.
party should bear the burden of uncertainty and risk.\footnote{See \textit{Jasanoff}, \textit{supra} note 23.} For example, a judge inclined towards a constructivist view of science might agree that there are multiple valid scientific positions on contested issues, rather than right and wrong answers. As such, she may be reluctant to exclude expert testimony. However, this heightened sense of scientific uncertainty might just as well lead a different constructivist-minded judge to become less liberal in his admissibility stance, as greater uncertainty justifies the admission of only extremely well-settled opinions. The decision of how one views science does not in itself answer the question of who should bear the risk of scientific uncertainty.

Third and most importantly, even if judges recognized, understood, and purported to adopt a particular epistemology of science, its impact on judicial decisions would still be limited. Current research on judicial behavior suggests that judges work to advance a variety of goals,\footnote{See \textit{Lawrence Baum}, \textit{The Puzzle of Judicial Behavior} (University of Michigan Press 1997).} such as enacting their ideological policy preferences,\footnote{See \textit{Segal} \& \textit{Spaeth}, \textit{supra} note 67.} engaging in strategic maneuvering to reach those policy preferences,\footnote{See \textit{Epstein} \& \textit{Knight}, \textit{supra} note 71.} advancing specific legal ideologies,\footnote{See \textit{Gillman}, \textit{supra} note 76.} navigating specific institutional concerns,\footnote{See \textit{Clayton} \& \textit{Gillman}, \textit{supra} note 71.} and following precedent.\footnote{See \textit{Richards} \& \textit{Kritzer}, \textit{supra} note 72; \textit{Thomas G. Hansford} \& \textit{James F. Spriggs II}, \textit{The Politics of Precedent on the U.S. Supreme Court} (Princeton University Press 2006).} Though scholars hotly debate the relative importance of these factors, it seems safe to say that maintaining a consistent epistemological theory would sit at the low end. I do not mean to argue that judges do not care about the quality of expert testimony or its impact on the trials in their courtroom; I simply imply that they will do so on a case by case basis, rather than working from a grand theory. Whether this occurs because judges are unaware of the theoretical basis for their decisions, because they deliberately avoid epistemological debates, or because they rationalize the use of a particular epistemology to achieve particular goals, pragmatism is the result.

Given these considerations, I see little chance for the emergence of a consistent epistemological understanding of \textit{Daubert}. Different areas of testimony will raise different arguments, bring forth different types of litigants, activate different sorts of judicial goals, and lead judges to be more receptive to specific realist or constructivist claims. Judges may rely on both realist indicators and constructivist concerns, depending upon the judicial goals in question and the type of expertise before them. Local constructions of reliability may indeed create shared understandings about how \textit{Daubert} should be carried out, in spheres as disparate as civil suits involving the drug Paradol, discussion over the validity of hedonic damages, or the reliability of handwriting expertise. However, such understandings will be constructed by trial judges \textit{sui generis}. Lacking clear doctrine, a clear stance on admissibility, and a clear stance on epistemology, \textit{Daubert} should allow judges to construct the reliability of particular types of evidence in accordance with a number of different variables, leading to what are in effect different admissibility standards for different types of evidence. This relative autonomy is further enhanced by the institutional context of \textit{Daubert} decisions, where
even when they occur, successful appeals are unlikely. As a result, we lack a universal conception of reliability, instead having local ones.

IV. EMPIRICALLY MODELING DAUBERT DECISIONS

A. Judges as Goal-Oriented Actors

If doctrinal analysis cannot predict admissibility decisions, must we simply compile these different constructions of reliability and eschew general theorizing? Perhaps. However, a more fruitful option would model Daubert decisions using the insights of the judicial behavior literature. That the application of Daubert itself cannot explain the variation among admissibility cases does not mean that the variation cannot be explained. Daubert analysis must join the manifold areas of law to which empirical study has been applied if we are to understand what drives the construction of reliability.

In this section, I suggest variables for building such a model. I ground the model in the well-regarded theory that views judges as goal-oriented actors in a particular institutional context. Judges are neither cogs in the mechanistic application of law nor unfettered policy-makers; like any political actor, they have goals that shape and are shaped by legal institutions (including the law). I suggest three categories of goals in admissibility decisions that might help identify relevant variables: (1) regulating the “reliability” of expert testimony, (2) maintaining institutional stability and autonomy, and (3) advancing one’s “policy preferences.” Beyond these three categories, I add the potential effect of on-point precedent, potential funding disparities among various classes of litigants, and specific circuit rules or practices that might shape admissibility decisions.

B. Legal Goals: Assessing Methodological Quality

First, I argue judges care about the quality, validity, or reliability of the expert evidence before them. Some commentators state that the first goal of courts is justice, not truth, and that an overly restrictive standard for expert testimony would harm the courts’ ability to resolve social disputes fairly. Oddly, this position dismisses the connection between finding truth and the perception of justice, that justice rests not only on fairness but also the belief that “correct” decisions have been reached. Downplaying truth in legal decisions could have ugly consequences: punishing an unsympathetic corporation on “social fairness” grounds could lead to actual harm if its product, was, in fact, safe and helpful. It may be true, of course, that in particular cases the uncertainty regarding the accurate distribution of blame cannot be reduced, and we will be stuck, as Jasanoff has argued, deciding on fairness grounds who should assume the burden of risk. Conceding this, I still assume most judges would agree that “getting it right” is an important part of their gatekeeping responsibilities, and that there will be some concern about the reliability and validity of the evidence before them.

137 See BAUM, supra note 131.
138 Farrell, supra note 21, 2205.
In line with Section III, I argue that the realist indicia from the *Daubert* dicta are too fluid to serve as independent variables. As the *Crisp* case illustrates, there are no fixed meanings of terms such as falsifiability, peer-review, or general acceptance. If a consensus on their meaning did arise within a particular circuit, or on a particular issue, they might serve as reasonably reliable variables, but such a consensus does not exist on a global level. I aim here for a parsimonious theory of judicial decision-making, and case-specific definitions of the dicta do not suffice. Moreover, as epistemological pragmatists, judges will not restrict their reliability analysis to the realist demarcations of good science found in *Daubert*, but may also consider “constructivist” insights—even if the judge is unlikely to understand them as such (or care about the difference). If the model is to include variables that measure judicial concern over reliability, something beyond the dicta must be included.

What factors might a generalist judge utilize as proxies for reliability in the aggregate? Without exhausting the possibilities, I suggest two variables here. First, faced with complex methodological disagreements, a judge might trust his instincts (or prejudices) and assume that particular classes of litigants are more likely to proffer expert testimony of dubious quality. This assumption fits with recent research suggesting that judges, like most people, often rely on intuition for their assessments rather than deliberation.\(^{140}\) I would argue that judges would be less likely to trust civil plaintiffs—whose experts they might see as nothing more than tools to advance the plaintiff’s chances of winning in the “big casino” of litigation—than civil defendants.\(^{141}\) This view of civil plaintiffs has been advanced at least since the 1990s by polemicists such as Peter Huber, who launched a war on “junk science” in the civil court system.\(^{142}\) Conversely, in criminal trials, judges might find defendants the more likely party to proffer unreliable expertise or baselessly attack experts proffered by the state, as their freedom depends on it.\(^{143}\)

Evidence that litigant identity impacts judicial decisions can be justified by more than just broad generalizations. Research by Clermont and Eisenberg, for example, demonstrates that federal appellate courts disparately favor civil defendants on appeal because they inaccurately view trial courts as biased for plaintiffs.\(^{144}\) The nascent empirical literature on *Daubert* also suggests admissibility decisions privilege civil defendants and the criminal prosecutors, although not for the same reasons. Civil cases are overwhelmingly levied by civil defendants against civil plaintiffs, with the latter losing the majority of these challenges.\(^{145}\) In the criminal sphere, by contrast, defendant

---

140 See Guthrie, *supra* note 11.
142 HUBER, *supra* note 18.
143 In other words, judges generally assume criminal defendants are guilty. See JAMES EISENSTEIN & HERBERT JACOB, FELONY JUSTICE: AN ORGANIZATIONAL ANALYSIS OF CRIMINAL COURTS (Little & Brown 1977); SAMUEL WALKER, SENSE AND NONSENSE ABOUT CRIME AND DRUGS: A POLICY GUIDE (Wadsworth Publishing 2000). That said, even if defense experts can be discounted because of their clients’ desperation, government experts are also susceptible to pressures to help the police and the prosecution secure convictions. See Michael J. Saks, *Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States*, 43 Sci. & Jus. 77 (2003).
145 See Risinger, *supra* note 8, at 102-11.
experts face the most challenges, even though prosecutors utilize the vast majority of experts.\footnote{Saks & Faigman, \textit{supra} note 10, at 120-25.} Like their civil plaintiff counterparts, criminal defendants are also far, far more likely to lose both their challenges to plaintiff experts as well as have their experts excluded than the reverse.\footnote{See Risinger, \textit{supra} note 8, at 110-111.} Even accounting for the reasonable objection that such disparities rest on the quality of the proffered expertise rather than distrust of particular litigants, litigant identity may serve as a rough and ready “constructivist” measure of reliability, one that should be tested.

A second proxy for reliability might come from positive or negative signals by professional groups or academic organizations. General acceptance remains part of the \textit{Daubert} regime, and such signals might anchor a judge’s considerations of whether an expert claim is sufficiently valid. Sociologists refer to attempts by academic or professional fields to demarcate the line between science and bad science or non-science as “boundary work.”\footnote{See, e.g., Thomas F. Gieryn, \textit{Boundary-work and the Demarcation of Science from Non-science}, 48 \textit{Am. Soc. Rev.} 781 (1983); Jasano\textsc{f}, \textit{supra} note 23.} The absence of a particular psychological theory from the DSM-IV,\footnote{\textit{American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders} (American Psychiatric Press, Inc. 2000).} for example, or a negative appraisal about a theory of medical causation from the American Medical Society, might shape the way in which a judge evaluates a particular expert. To cite a particular example, a number of toxic tort plaintiffs have brought forward experts to claim that a company’s product has led to them to suffer from Multiple Chemical Sensitivity (MCS), a syndrome which one commentator calls a “multi-symptomatic disorder affecting multiple organ systems resulting from exposure to a multitude of chemicals at levels tolerated by the majority of the population.”\footnote{Carl H. Johnson, \textit{When Science is too Daunting: Multiple Chemical Sensitivity, Federal Courts, and the Struggling Spirit of Daubert}, 11 \textit{Vill. Envtl. L.J.} 273, 276 (2000).} To the mainstream medical world, MCS is fatally lacking in objective diagnostic criteria injury that might permit a doctor to know when a patient does or does not suffer from it.\footnote{See William J. Waddell, \textit{The Science of Toxicology and its Relevance to MCS}, 18 \textit{Reg. Toxicology \& Pharmacology} 13 (1993).} Courts have been quick to cite major medical organizations when they lambast the adherents of MCS (usually called clinical ecologists) as promulgating unscientific theories, sometime explicitly relying on such statements to exclude the plaintiff’s experts.\footnote{See, e.g., Frank v. New York, 972 F. Supp. 130, 133-34 (N.D.N.Y. 1997); Treadwell v. Dow United, 970 F. Supp. 974, 981 (M.D. Ala. 1997); Coffin v. Orkin Exterminating Company, 20 F. Supp. 2d 107, 111 n.5 (D. Me. 1998).}  

\section*{C. Institutional Goals: Weighing the Effects of Admissibility Decisions on Courts}

While reliability may be the primary consideration in admissibility decisions, it is certainly not the only one. As “repeat players,” judges should see their decisions as iterations in a larger process, rather than a series of unconnected outcomes.\footnote{Galanter, \textit{supra} note 26.} Decisions which may impact the operation of court institutions, then, should garner special attention. The Supreme Court’s recent decision in \textit{Blakely v. Washington}, for example,
which held the Sixth Amendment to require all facts introduced at sentencing to have been heard by a jury, was criticized by Justice O'Connor as much on institutional grounds as constitutional ones.\textsuperscript{154} O'Connor worried that the decision would impose a “constitutional tax” on jurisdictions that wished to use facts at sentencing that had not been brought forward during the guilt phase, requiring them to employ costly bifurcated trials for crimes short of murder.\textsuperscript{155} Ostensibly constitutional concerns, such as standing or justicibility in general, may rightly be seen as guided by concerns over the courts’ institutional capabilities.

In \textit{Daubert} decisions, judges might consider how their decisions to admit or exclude certain types of expert testimony might harm the courts. Admitting certain types of plaintiff expert testimony in the civil sphere might signal other litigants to file similar suits, increasing caseload pressures or greatly complicating previously routine decisions. Conversely, excluding prosecutorial expert testimony might encourage criminal defendants to escalate their own use of \textit{Daubert}. The \textit{Crisp} majority, for example, argued against the “wholesale exclusion of a long-accepted form of expert evidence,” a phrase that implies a fear of institutional consequences for granting the \textit{Daubert} motion to exclude.\textsuperscript{156} Other admissibility decisions have been more direct in their consideration of institutional impact. In response to a \textit{Daubert} motion to exclude fingerprint expertise on grounds similar to those made by Crisp, the judge worried that:

\begin{quote}
This Court is unpersuaded by such reasoning, for it hearkens to an imprudently stringent understanding of scientific objectivity… Indeed, such reasoning could function to render numerous categories of expert evidence, such as psychiatric or medical forensic evidence which rest in some part upon that individual’s skill and experience in analyzing data, unreliable. \textit{Daubert} and its progeny simply do not mandate such a conclusion.\textsuperscript{157}
\end{quote}

A similarly situated judge in \textit{United States v. Hilerdiue-Alteme} makes an even more direct reference to institutional concerns:

\begin{quote}
The difficulty with the argument advanced by the defenses is that it proves too much. If Dr. Story’s criticism of the process of fingerprint examination were sufficient to preclude the testimony of other experts, large categories of scientific and technical testimony would be inadmissible. At a minimum, it would be necessary to eliminate the defense of insanity, since virtually all psychiatric opinions are subjective in whole or in part.\textsuperscript{158}
\end{quote}

Note that the latter opinion essentially bases its application of \textit{Daubert} upon the institutional consequences of exclusion, rather than addressing reliability per se. Given

\textsuperscript{155} Id. at 317.
\textsuperscript{156} United States v. Crisp, 324 F.3d 261, 268 (4th Cir. 2003).
\textsuperscript{158} United States v. Hilerdiue-Alteme, 99-8131 CR, at *7 (S.D. Fla. April 7, 2000).
these examples, consideration of how the admissibility decision might impact caseload or bring new complexity to cases should be included in the behavior model.\textsuperscript{159}

Institutional variables should not be restricted to caseload. A more nuanced way in which categories of expert testimony might “threaten” courts is when particular expert theories challenge the empirical assumptions on which courts operate. For example, the act of witnessing—the presentation of evidence by laymen regarding things they have seen—has been a pillar of our trial system for centuries. Unfortunately, psychologists who study the ways in which individuals process, store, and recall memories convincingly contend that eyewitness testimony is all too fallible.\textsuperscript{160} The Innocence Project, famous for freeing wrongfully convicted defendants using DNA evidence, estimates on its website that mistaken eyewitness identification played at least a contributing role “in more than 75\% of convictions overturned through DNA testing.”\textsuperscript{161} Unsurprisingly, when the case against a criminal defendant is based wholly or in large part upon eyewitness testimony, some defendants have attempted to present expert testimony regarding these problems.\textsuperscript{162}

Under either a liberal or a stringent admissibility standard, expert testimony on eyewitnesses should be reliable.\textsuperscript{163} The field’s array of claims has been tested and peer-reviewed multiple times, and there even exists concrete data regarding the degree of general acceptance of particular claims among experts.\textsuperscript{164} Despite these bona fides, however, eyewitness experts have often met with a frosty reception from federal judges during admissibility hearings. Appeals regarding the exclusion of eyewitness testimony, for example, have found the appellate judges offering vague or unusual arguments when finding no abuse of discretion in the decision of the trial court.\textsuperscript{165}

\begin{footnotesize}
\textsuperscript{159} Such a variable is also needed to help tease out the true meaning of the litigant identity variable. For example, do civil plaintiffs and criminal defendants generally lose more from Daubert decisions because judges perceived their experts to be of lower quality than other litigants, or because deciding in their favor might substantially increase judicial workloads? Or, perhaps, do these variables interact?
\textsuperscript{160} CUTLER & PENROD, supra note 25; LOFTUS, supra note 25.
\textsuperscript{162} See CUTLER & PENROD, supra note 25. These limits include how an individual’s ability to remember a face accurately declines rapidly after long periods of time, the weakened ability to remember a face when a weapon is present, or the additional difficulties in eyewitness recall that occur when the witness is of a different race than the person he witnesses.
\textsuperscript{163} Some critics, such as Ebbe Ebbesen, have argued that because many of these claims have been tested in mock “crimes” on university campuses, they lack sufficient external validity. See, e.g., United States v. Burton, 1998 U.S. Dist. LEXIS 18730, at *36-43 (E.D. Tenn. October 8, 1998).
\textsuperscript{165} United States v. Amador-Galvan, 1997 U.S. App. LEXIS 5320 at *4-5 (9th Cir. March 14, 1997). The judge in Amador-Galvan upheld the exclusion of an eyewitness expert on the grounds that “the testimony is based upon suspect, scientific analysis in that the alleged scientific analysis is abstract, incomplete, and disregards many known variables impacting upon the receipt of evidence and the decision making process of juries.” What these other variables are, and why the evidence is incomplete, is unclear from the opinion, though in fairness, an inadequate Daubert proffer may have been to blame. The judge in United States v. Brien, by contrast, upheld the district court’s exclusion by noting that the expert’s testimony “does not concern a single long-established scientific principle such as whether radar can measure speed.” United
published admissibility decisions regarding eyewitness experts, however, avoid reliability concerns altogether and focus on the institutional problems that eyewitness experts may create. Despite a 1984 amendment to Federal Rule of Evidence 704 that states, “testimony in the form of an opinion or inference otherwise admissible is not objectionable because it embraces an ultimate issue to be decided by the trier of fact,” some decisions openly worry that eyewitness experts usurp the role of the jury by “deciding” witness credibility. In his concurring opinion in United States v. Hall, Judge Easterbrook suggests the basis for this fear:

Consider, for example, the proposition—fundamental to any system that relies on lay adjudicators—that jurors understand and follow the instructions given by the court. It may be that jurors don't understand legalese, or if they understand the instructions don't follow them. Or consider cross-examination. Jurors may believe that witnesses who hesitate, perspire, or fidget during cross-examination are hiding the truth. This is the view that underlies polygraph examinations, but without the precision of measurement...Because trials rest on so many contestable empirical propositions, including those about eyewitness recollection, it always would be possible to offer expert evidence along these and related lines.

“A trial about the process of trials,” he later contends, only distracts the jury from the issue of the criminal charge at hand. In other words, the admission of expert eyewitness testimony or polygraph experts into court might have long-term deleterious effects on trials, particularly if their admission encourages defendants to use social science research to challenge other foundations of the trial process. The possibility that courts will be particularly unreceptive to such testimony merits some attention in empirical work on admissibility decisions.

D. Attitudinal Goals: Advancing Judicial Policy Preferences

Judicial goals also may include policy preferences, the desire to reach particular outcomes or patterns of outcomes regardless of the law. Judicial attitudes will be connected to the larger political framework by the process of appointment; judges appointed by conservative presidents may have different desired legal outcomes than those appointed by liberal executives. While judges are not legislators, they do retain considerable freedom to shape decision outcomes, and often do so along a standard

---

166 FED. R. EVID. 704.
167 United States v. Hall, 165 F. 3d 1095, 1119 (7th Cir. 1999).
168 Id. at 1119-20.
169 Simmons, supra note 3.
170 One such (pre-Daubert) example noted by Easterbrook is Gacy v. Welborn, 994 F. 2d 305 (7th Cir. 1993), in which the defendant appealed his death sentence, using social science research to attack the clarity of the court’s jury instructions on mitigating factors.
liberal-conservative continuum. As discussed in Section III, such discretion varies among different judicial contexts—a Supreme Court justice ruling on an issue of constitutional law has far more discretion than a trial court judge implementing a matter of routine procedure. *Daubert* decisions, I have noted, are contexts where trial court judges are relatively unconstrained, and there is little reason to expect that liberal and conservative positions on torts and criminal prosecution will be absent from admissibility decisions. A more conservative judge, for example, might tend to side with business litigants in tort cases, while a liberal judge might be more apt to question the government’s use of expert testimony in criminal cases.

Some measure of judicial ideology must also be included as a control for previously discussed factors. Do all judges view civil plaintiffs as more likely to offer unreliable testimony, or are only conservative judges more likely to do so? Would a liberal judge be more willing to tempt caseload pressures by limiting or excluding prosecutorial testimony? Such findings might show that a construction of reliability was heavily mediated by prior political positions, rather than a more widespread, institutional suspicion of evidence offered by particular litigant classes. While mapping standard liberal-conservative attributes onto federal judges is not an easy task, political scientists who study judicial behavior have developed reasonably valid measurements for doing so.

**E. External Variables**

Finally, some variables in the admissibility process will be independent of the deciding judge. I have already suggested that despite their relative autonomy, district judges may constrain themselves to on-point precedent, whether because of normative concerns or to reduce workload. The proper normative role of precedent in judicial decision-making, as well as its actual empirical impact, is hotly debated by both scholars of judicial behavior and the justices of the Supreme Court. Where precedent is surely at its strongest, however, is when the facts of that precedent and the facts in the current case essentially mirror each other. This might occur with some regularity in

---

171 See Segal & Spaeth, supra note 67.
172 See Buchman, supra note 7, at 682-87.
173 The best such measurement, in my view, is the Giles score. Creating a Giles score begins by relying on Poole’s DW-Nominate scores scores (for each Senator and President, which vary from -1 to 1 on a liberal-conservative scale. Federal judges receive the appointing President’s ideology score, unless their home-state senators are the same party as the President. In this case, the judge is either given either the home-state senator’s Poole score, or if both home-state senators are of the president’s party, the average of the two. This method has two advantages over simply coding federal judges as appointed by Republicans or Democrats. First, it recognizes that political party and attitudes are related but distinct; G. W. Bush appointees are more likely to be conservative than Ford or Eisenhower appointees. Secondly, it accounts for the potential moderating influence of home-state senators; Nixon, for example, may have appointed more liberal justices in Pennsylvania than in Utah, because the Republican senators in Pennsylvania were more liberal than he was.
174 See, e.g., Brishin, supra note 70; Gillman, supra note 80; HANSFORD & SPRIGGS, supra note 136; Jeffrey A. Segal & Harold J. Spaeth, The Influence of Stare Decisis on the Votes of United States Supreme Court Justices, 40 AM. POL. SCI. REV. 971 (1997).
176 See Richard L. Pacelle, Jr et. al., Keepers of the Covenant or Platonic Guardians? Decision Making on the U.S. Supreme Court, 35 AM. POL. RES. 694 (2007).
Daubert cases, where individual litigants in different courts move to exclude or introduce similar types of expert testimony. Given the factors discussed above, there is likely to be some variation in how different judges address a particular type of expert testimony initially; as time goes on, however, a particular construction of reliability may become dominant. Decisions that are unusually thorough, thoughtful, or time-consuming should be more likely to become precedent, especially since a judge should be less likely to duplicate an already existing Daubert analysis. The greatest opportunity for the full play of reliability concerns, institutional concerns, and policy preferences to shape a particular construction of reliability may be in the first few instances a particularly type of expert testimony faces admissibility hearings. Afterwards, the costs of engaging in full-bore Daubert analyses may be too high, and variation will tend to be greater between different types of expert testimony rather than within them.

Second, the model should consider the relative resources that individual litigant classes bring to the table. A judge might grant the prosecution’s motion to exclude a psychologist who would testify on eyewitness testimony not because he assumes the defendant is desperate or fears the effects of such witnesses on the court system, but because the defense attorney has failed to describe what that expert will do, why he satisfies Daubert, or why his testimony will be reliable, relevant, and helpful. The success rate for civil defendants, for example, may in part be explained by their superior resources in hiring high-quality experts and retaining attorneys with considerable experience in dealing with admissibility issues. An accused criminal, by contrast, may rely on a public defender with little or no experience in writing Daubert proffers, or have insufficient funds to hire an expert with adequate credentials. Litigant resources should also be included in the model to better isolate the possibility that judges use litigant identity as a proxy for reliability, as discussed above, or simply dislike particular litigants on policy grounds.

Next, judges may have background assumptions about the reliability of certain fields or types of testimony, or cultural assumptions that dispose them to view these fields in a favorable or unfavorable light. Jennifer Mnookin, for example, has convincingly argued that the legal system of the late 19th and early 20th century had inculcated the belief that individuals generally and criminals specifically possessed unique physiological traits that permitted “comprehension of a deeper, otherwise unattainable reality.” As such, judges were more inclined to accept various nascent forensic individuation techniques as valid and reliable. A better exploration of current paradigms might reveal similar tendencies of courts and society alike to view more readily certain ideas as scientific.

---

177 My own sense of such briefs, though I have read only a few, is that they vary wildly in quality; experience with Daubert motions and a substantive understanding of the underlying science are the most likely causes of this variation. I do not expect that these qualities will be distributed evenly throughout all categories of attorneys.


Finally, it is possible that circuits themselves serve as relevant independent variables. Circuits may differ in their interpretation and application of Daubert, leading to differences that do not draw the Supreme Court’s attention but may affect outcomes nevertheless.180 These rules and norms may constrain judicial behavior in ways that Daubert itself does not do, and are worth including as dummy variables.

V. CONCLUSION: NOTES ON GETTING DAUBERT DATA

Building and testing such a model is beyond the scope of this paper. I hope, however, I have made a convincing case that doing so is useful and necessary. Empirical research on Daubert has already begun: Risinger, for example, has systematically examined differences in decision outcomes between civil and criminal admissibility decisions,181 while Buchman has found that political ideology affects admissibility decisions in federal district court decisions.182 The body of work is sparse, however, and many questions remain. Does presumed reliability, superior resources, or political attitudes lie behind the significant advantages which accrue to criminal prosecutors and civil defendants? When do institutional concerns affect Daubert decisions, and how can one measure the effect of those concerns? Are local constructions of reliability for new technologies heavily path-dependent, making initial or particularly studious judicial assessments of reliability the critical cases? I hope my article has made the case that such questions are interesting and necessary.

I conclude by noting one major obstacle to the empirical study of Daubert. A primary problem, for either qualitative or quantitative studies, is ensuring that inferences are drawn from representative data. However, obtaining a representative or random sample of admissibility decisions from the universe of all federal district court decisions will be quite difficult. Standard database searches, such as WestLaw and Lexis, garner some district court decisions, but because the available cases are a small and possibly unrepresentative sample of the whole, relying on them for generalizations risks serious inferential errors. If, for example, decisions to exclude expert testimony are more likely to be published, one might overestimate the degree to which all decisions are excluded if only published cases are studied.183 Relying on appellate decisions for data mostly overcomes this problem, as federal appellate decisions, at least new ones, are reliably recorded. The gain is illusory, though, as the benefit of easier sampling is offset by the skewed picture of Daubert decisions appellate opinions create. A more valid study of

181 Risinger, supra note 8.
182 Buchman, supra note 7.
183 Peter Siegelman & John J. Donohue III., Studying the Iceberg from its Tip?: A Comparison of Published and Unpublished Employment Discrimination Cases, 24 LAW & SOC'Y REV. 1133 (1990). There is another problem, of course, with relying on judicial decisions, namely ignoring the possibility that Daubert’s largest impact is on persuading parties to settle, rather than on trials. I concede this probability; I simply wish to study a different facet of the process. For a look at how Daubert factors in to the early stages of litigation, see Herbert M. Kritzer, Daubert in the Law Office: Routinizing Procedural Change, 5 J. EMPIRICAL LEGAL STUD. 109 (2006).
Daubert decisions must be conducted “on the ground,” where the full range of issues are addressed, and the judges do not employ Joiner in reviewing a previous determination.

In other research agendas, scholars have relied upon case data from the Administrative Office of the United States Courts to create a representative data base.\footnote{Clermont & Eisenberg, supra note 144; Kevin M. Clermont & Theodore Eisenberg, Appeal from Jury or Judge Trial: Defendants' Advantage, 3 AM. L. & ECON. REV. 125 (2001).} The Federal Judiciary’s PACER (Public Access to Court Electronic Records) system may also grant researchers access to the closet approximation of the actual universe of district court cases, at least for recent years.\footnote{See http://www.pacer.psc.uscourts.gov. Most, but not all federal jurisdictions make their work product available for PACER.} However, these free or low-cost solutions suffer from a lack of appropriate search capabilities. Though quite useful for studies that test general propositions about federal decisions, these databases do not permit the ready retrieval of Daubert decisions. Emerging commercial solutions such as Daubert Tracker,\footnote{See http://www.dauberttracker.com.} LexisNexis CourtLink,\footnote{See http://www.courtexpress.com.} or WestDockets\footnote{See http://west.thomson.com/documentation/westlaw/wlawdoc/wlres/litwd4b.pdf.} rely on the same information as PACER but may provide the search capabilities that make a systematic study of Daubert district decisions feasible.\footnote{For more information on these options, see Linda Fields, Using Docket and E-Filing Retrieval Systems for Legal Research, 33 THE COL. L. AW. 61 (2004).} That said, such solutions may be too expensive for the individual researcher to employ. Moreover, broader searches such as these will require limiting the study to a few districts or a narrow slice of time, barring prodigious resources. Regardless of how the researcher ultimately selects cases, she must understand the tradeoffs inherent in that choice and adjust her interpretation of the data and her presentation of the results accordingly.

None of this is to say that empirical research on Daubert must be quantitative. Indeed, while quantitative research provides more reliable generalizations for large numbers of cases and a more precise estimate of uncertainty, qualitative explorations of particular case areas may uncover new variables, explicate the relationship between them, or give the researcher a better understanding of the judge’s purported justification for the decision. This is particularly true for the institutional concerns listed in Part IV, which may not easily translate into quantitative measures. In either case, if scholars and practitioners alike want a better understanding of admissibility decisions, as well as a reasonable shot at accurately predicting judicial reactions to new technology like fMRI scans, there is much work to do.