

***Daubert* Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers**

*Christopher B. Mueller**

Daubert is one of the more important decisions of the twentieth century because it changed fundamentally the relationship between law and science.¹ Prior to *Daubert*, the law deferred to the scientific community on the question whether answers that scientists provide are sufficiently grounded in theory and practice to be trusted and acted upon by courts. After *Daubert*, judges are charged independently to appraise what science has to offer, in effect screening out evidence offered as science if it is invalid or unreliable.

To put it another way, a pre-*Daubert* judge who might have hesitated to exclude what seemed to be testimony on a matter of science could say, in effect, “it is not the court who rejects what you say, but other experts in your field.” A judge fearful of criticism for *admitting* such testimony could say, in effect, “it is not the court who endorses what this expert has to say, but credentialed people in a recognized discipline.” A post-*Daubert* judge has less room to hide. If he excludes evidence proffered as science he is expected to say “the court finds that what you say is not sufficiently grounded in theory or

* Henry S. Lindsley Professor of Law, University of Colorado School of Law. I wish to thank participants in the Seton Hall Symposium on Expert Admissibility in February 2003 who made comments and suggestions that helped me in revising this article. They include Professors Ron Allen, David Barnes, Margaret Berger, Neil Cohen, David Faigman, Richard Friedman, Edward Imwinkelried, Roger Park, Michael Saks, and Joseph Sanders. I also want to thank Professors Mark Denbeaux and Michael Risinger for organizing the Symposium, Professor David Bernstein for commenting on a draft of this article, and Ryan Philp, Joe Arnold, and John Falzone of the *Seton Hall Law Review* for their hard work on the Symposium and the articles published in this issue. I also thank Kevin Doyle for proofreading and editing the manuscript of this article with great care and sensitivity. Needless to say, I remain responsible for whatever errors may appear in this piece.

¹ *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 588 (1993) (noting that FRE 702 superseded the *Frye* standard; under FRE 702, scientific evidence must be valid in the sense of being reliable and must “fit” the case and, even if the evidence does satisfy these requirements, it is subject to the possibility of exclusion under FRE 403) (citing *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (adopting what has come to be known as the “general acceptance” standard)).

practice,” or “lacks sufficient basis in fact” or “lacks sufficient connection to the case at hand.” A post-*Daubert* judge who admits such evidence is expected to say “the court finds that indeed this testimony is properly grounded in theory and practice, and adequately based on the facts and sufficiently related to the task at hand.” To be sure, *Daubert* still leaves room to hide. Factors like “peer review”² and “general acceptance” provide opportunities, as does the possibility of invoking FRE 403, and a judge can also distance himself by casting his decision in terms of “adequate assurances” or “inadequate assurances” of validity.³

The basic point, however, is that *Daubert* puts judges into the position of judging science. That makes *Daubert* revolutionary.⁴

Criticisms of *Daubert* abound, particularly in toxic tort cases. Perhaps such criticisms are inevitable when a single case so profoundly changes the legal landscape. The most serious criticisms are advanced from three perspectives: One is epistemological and structural (or “political” in the fine sense of the term). This criticism holds that judges are not much more able than juries to appraise proof offered as science, and that attempts to exercise the “gatekeeping” role infringe on the powers and responsibilities of juries to act as factfinders. Another criticism is pragmatic, substantive, and to some extent ideological. This criticism holds that judges applying *Daubert* are throwing out too much good evidence proffered by civil claimants in toxic tort cases. A third is philosophical. This criticism holds that *Daubert* misunderstands science.

What follows is a defense of *Daubert* against these criticisms, followed by a suggestion of my own, which is that the *Daubert* revolution would achieve more if appellate courts abandoned the

² For an argument that “peer review” is not what it is cracked up to be, see Joelle Anne Moreno, *Eyes Wide Shut: Hidden Problems and Future Consequences of the Fact-Based Reliability Standard*, 34 SETON HALL L. REV. __ (upcoming in Fall 2003).

³ See generally Michael H. Graham, *The Expert Witness Predicament: Determining “Reliable” Under the Gatekeeping Test of Daubert, Kumho, and Proposed Amended Rule 702 of the Federal Rules of Evidence*, 54 U. MIAMI L. REV. 317, 317 (2001) (arguing that judges should not determine “whether the explanative theory actually works” to produce an accurate conclusion, but “whether there are sufficient assurances” that it does).

⁴ One crude measure of *Daubert*’s impact can be seen just by glancing at citation history. In the 38 years between 1945 and the decision in *Daubert* in 1993, *Frye* was cited in approximately 260 reported federal cases and 800 reported state cases. As of this writing, *Daubert* is almost ten years old. As of January 25, 2003, *Daubert* has been cited in nearly 2,500 reported federal cases and 1,200 reported state cases, while *Frye* has been cited in approximately 270 reported federal cases and 1,000 reported state cases.

abuse-of-discretion standard in reviewing the rulings of trial judges in this area. To be fair, I should note that most critics have not called for the abandonment of *Daubert* and few would endorse a return to the *Frye* standard. Casting the criticisms in their best light, their aim is to improve *Daubert*, an undertaking that I gladly join. This essay addresses civil rather than criminal cases,⁵ and scientific evidence rather than “experiential expertise,” even though this dichotomy is hard to draw and counts for less than it once did because the *Daubert* standard applies to all expertise.⁶

I. JUDGES CAN DO BETTER THAN JURIES: *DAUBERT* GATEKEEPING MADE REAL

In a nutshell, *Daubert* is the right standard because it asks directly the question that *Frye* put only indirectly, and thus puts courts in a better position to arrive at satisfactory answers. The central issue is scientific “validity,” and the criteria suggested by *Daubert* are useful in resolving that issue.

Here it is worth pausing to ask some pragmatic questions: Why have a validity standard to begin with? Why not simply approach science with the kind of openness suggested by FRE 702 on its face? In other words, why not simply admit scientific evidence if it seems relevant and helpful and the witness is qualified?

The answer given by the Court is more positivist than policy-based. FRE 702 *requires* science to satisfy a validity standard, so courts are bound to scrutinize such proof. That answer is unsatisfactory because it does not emerge from the “plain meaning” of FRE 702 or even a reasonable interpretation of the Rule’s language. The Court has acknowledged that the Rules did not displace all prior evidence

⁵ I’ve heard enough from able commentators, including Professor Risinger, to be convinced that courts are too credulous with purported scientific evidence in criminal cases. *Daubert* seems to have exposed the soft underbelly of forensic science, and I think that judges, to the extent they see the problem, are troubled more by the prospect that the system will break down than by the prospect that unreliable evidence is being used to convict. There is such a thing as too much revolution, and the impact of *Daubert* in criminal cases has yet to be worked out in a satisfactory way.

⁶ See *Kumho Tire v. Carmichael*, 526 U.S. 137, 141 (1999) (“*Daubert*’s general holding . . . applies not only to testimony based on ‘scientific’ knowledge, but also to testimony based on ‘technical’ and ‘other specialized’ knowledge” under FRE 702). For an effort to develop standards, consistent with *Kumho* and *Daubert*, by which the validity of nonscientific expertise might be appraised, see Edward J. Imwinkelried, *The Next Step After Daubert: Developing a Similarly Epistemological Approach to Ensuring the Reliability of Nonscientific Expert Testimony*, 15 *CARDOZO L. REV.* 2271, 2293 (1994) (stating that a trial judge can exclude experiential nonscientific expertise when it is based on no experience or only limited experience, and when the experience is too dissimilar from the issues at hand).

doctrine, and this point is critical to another major holding.⁷ The conclusion in *Daubert* rests on the notion that the word “scientific” as used in FRE 702 is a rich or deep normative term that implies a standard of legitimacy. It is of course astonishing, if we suppose that this meaning really is to be found in FRE 702, that nothing in legislative background supports this reading (in fact the term seems merely descriptive). In truth, the Rules provide no compelling basis for discarding the old *Frye* standard. What we now call the *Daubert* standard is in reality judge-made law disguised as something else. That is not to say that I disapprove of the decision, for the opposite is true: I think *Daubert* represents an advance, that it is at least consistent with the elastic contours of FRE 702, and that it good lawmaking, even if disingenuous in its logic.⁸

But there are questions that should be asked before reaching that conclusion: Should we have such a standard? Does it make the law better? Keeping the focus on civil cases, my answer is yes.

I think that three facts of modern life conspire to suggest that we need a validity standard: First, we ask courts to resolve difficult technical and scientific issues. Second, much scientific knowledge is fluid and contestable, inaccessible to laypeople, hard to understand, and qualified in ways that elude ordinary experience and intuitions. Third, our adversary system places primary responsibility for gathering and presenting evidence in the hands of the parties, and creates incentives that lead to risks.

What we see is something like this: Complex questions arise, which can be answered, if at all, only by calling on scientific expertise. A salient modern example is the question of causation in the toxic tort setting. Parties and courts look to science for the answer. Lawyers on each side find experts who agree to help for a price. The issue is joined, and we discover that there is no definitive answer and only partial information—studies and theories, usually involving

⁷ See *United States v. Abel*, 469 U.S. 45, 50 (1984) (explaining that the drafters of the Rules could not have intended to “scuttle entirely” the practice of cross-examining to show bias, even though they offer no “express treatment” of the subject); see also Edward Cleary, *Preliminary Notes on Reading the Rules of Evidence*, 57 NEB. L. REV. 908, 915 (1978) (stating that after adoption of the Rules, “no common law of evidence remains,” at least “in principle,” but “in reality” the situation is different, because “the body of common law knowledge continues to exist” in the “somewhat altered form of a source of guidance in the exercise of delegated powers”) (internal citations omitted).

⁸ See generally RONALD DWORKIN, *LAW’S EMPIRE* 255 (1986) (contrasting, *inter alia*, law as pragmatism with law as integrity, meaning that judges seek to resolve hard cases by some “coherent set of principles” in order to make the “complex structure” of law and politics “the best these can be”).

some combination of chemical structure analysis, animal tests, epidemiology, and/or “differential diagnosis.” None of the proof either does or can answer *directly* the question of individual causation (“specific” causation). Instead, such proof shows a *possibility*, and perhaps sometimes a *probability*, of causation in an individual case by suggesting that the substance in question *can* or *does* cause some ailments in some people (“general” causation) or by suggesting that no *other* explanation is likely (differential diagnosis testimony).

What can courts reasonably do? Here is one possibility: Courts can suppose that the data and conclusions presented by qualified experts reflect valid science, upon which our system can reasonably allow a jury to rely in rendering a verdict for or against recovery in some very substantial amount. That seems close to the view in jurisdictions that admit scientific evidence on the basis of a credentials test coupled with findings that the proof is relevant and helpful.⁹ But there is another possibility, which seems more realistic: We can make the judgment that not all evidence that is presented as science, even by qualified witnesses, is of such quality that it can be relied upon to make serious decisions of the sort required for civil judgments. We can believe that such evidence varies in quality, and that sometimes it is *not* reliable enough. We can suppose that gaps in scientific understanding create room for interpretive disagreement, and that financial incentives, whether arising from the involvement of scientists in commercial or other funded projects or from their involvement in litigation, can compromise expert testimony. We can believe that science, like law, leaves room for principled intellectual disagreement that reflects differences in technical understanding or personal philosophy. We can also suppose that these differences sometimes lead to errors or to conclusions that cannot be defended or would be condemned by most others of similar training. Obviously, *Daubert* reflects the latter view of science, and I think that is the more realistic view.

There is yet another question. Should we charge judges to be

⁹ See, e.g., *State v. Peters*, 534 N.W.2d 867, 871-73 (Wis. Ct. App. 1995) (following neither *Daubert* nor *Frye*, in applying WRE 702, the trial judge is limited to considerations of relevancy, qualifications of the witness, whether the evidence is superfluous or will waste time or resources, whether probative value is outweighed by prejudice, whether jury can draw its own conclusions, whether the evidence is inherently improbable, and whether the area is suitable for expert opinion); *Green v. Smith & Nephew AHP, Inc.*, 617 N.W.2d 881, 890 (Wis. Ct. App. 2000) (finding that unlike *Daubert* jurisdictions, where “the trial court has a significant ‘gatekeeper’ function in keeping from the jury expert testimony that is not reliable, the trial court’s gatekeeper role in [this state] is extremely limited”).

the gatekeepers, or should we fold the gatekeeping responsibility into the factfinding responsibility, leaving the assessment of science to juries? As others have suggested, the right approach is to ask this question: Are judges more capable than juries of playing this role? In his engaging contribution to this conversation, Professor Joseph Sanders says yes.¹⁰ His conclusion rests on an examination of empirical data (some published, some new and unpublished), and he is cautious. Still, his conclusion is generally yes.

Professor Sanders builds on what he calls a “counter-revolutionary” Kansas decision and the work of Professor Alvin Goldman.¹¹ Goldman suggests that we should consider (a) the characteristics of the audience (juries), (b) the characteristics of the witnesses (experts, including scientists), (c) the criterion to be applied (*Daubert* or some other standard), and (d) other alternatives. As for juries, we have indications that they have trouble with complex cases, and with scientific evidence, and we have reason to believe that better-educated juries do better in these areas. We have indications that juries approach expertise with skepticism. We have indications that juries appraise expert testimony not by grappling with technical issues, but by counting extraneous factors like qualifications, the *number* of arguments (rather than quality), and personal attractiveness. We understand that jurors give more credence to messages framed in simple language, less to those framed in complex language, and they pay close attention to demeanor. As for experts, we have confirmation of what we have long suspected: They tailor their testimony to please whoever pays them. They learn to “perform” in court. As for judges, we have some mixed news: Data on state judges suggest that many do not understand the “testability” concept (can the evidence be “falsified”?) or error rates, although

¹⁰ Joseph Sanders, *The Paternalistic Justification for Restrictions on the Admissibility of Expert Evidence*, 33 Seton Hall L. Rev. 881, 937-38 (2003); see also Brian Leiter, *The Epistemology of Admissibility: Why Even Good Philosophy of Science Would not Make for Good Philosophy of Evidence*, 1997 BYU L. REV. 803, 814-15 (1997) (referring to rules designed to substitute “the rulemaker’s judgment about what is epistemically best for agents for their own judgment” as “epistemic paternalism”).

¹¹ The Kansas decision is *Kuhn v. Pharm. Corp.*, 14 P.3d 1170, 1173-74 (Kan. 2000) (refusing to apply the state’s *Frye* standard to testimony by treating physician, based on differential diagnosis, that Parlodel caused a new mother to suffer stroke) (Parlodel is a drug taken by mothers, who prefer not to breast feed their babies, in order to suppress lactation). The court’s reasoning in *Kuhn* involves consideration of factors similar to those advocated by Professor Goldman, except that the court here decides that those factors support the conclusion that juries can appropriately evaluate such testimony, so the judge need not play screening role. The main work by Professor Goldman on which Professor Sanders draws is Alvin Goldman, *Epistemic Paternalism: Communication Control in Law and Society*, 88 J. of Philosophy 113 (1991).

they do better with criteria of peer review and general acceptance. Surveys of federal opinions, however, suggest that judges are achieving a better understanding of science. As for alternatives, we have some indications that cross-examination does little to affect jury appraisals of expert testimony.

Looking at this data, Professor Sanders concludes that they provide “some support” for restricting admissibility by a standard applied by judges. He comments as well that his own reading of the cases indicates that courts are doing better than the survey of state judges suggests. I agree, and I too can report that reading many decisions leads me to believe that appellate judges are doing better in appraising scientific proof than they did in the days of *Frye*.

At the risk of being simplistic, I think four additional factors point toward the need for a validity standard in which trial judges screen out questionable science. First, on balance judges are better educated than juries and are selected with attention to merit and skill. Second, judges have experience with adversarial presentations and are likely to be better able to understand the substance of testimony and its relationship to the issues. Both of these points suggest that judges can do better than juries in separating what should count from what should not. Third, the complexity of scientific evidence suggests that the “relevancy” criterion that applies to other evidence is not adequate to deal with science. Although I cannot prove it, I suspect that the very fact that a court *admits* evidence that is daunting or complex conveys to jurors an unspoken message of invitation, suggesting that they *can* rely on it (even though they need not).¹² Finally, in jury-tried civil cases, it seems wiser to have judges decide the validity point simply because it is better to separate the decision on this point from the decision on the merits. The point is not merely that “two heads are better than one” (counting one head for the judge and one for the jury), but that one can reasonably expect a better decision on validity by someone who is not *also* responsible to decide whether the plaintiff or the defendant has the stronger case.¹³

¹² That jurors’ expectations may affect the way they process the evidence they hear is recognized in a different context in *Old Chief v. United States*, 519 U.S. 172, 188-89 (1997) (stating that juror expectations may arise “from the experience of a trial itself,” and that shifting from descriptions by witnesses “naturally” describing “a train of events” to a different kind of presentation may make jurors “wonder what they are being kept from knowing”).

¹³ This point was raised during the Seton Hall Symposium, *Expert Admissibility: Keeping Gates, Goals and Promises*, in February 2003, but none of the various participants, whom I contacted, claims credit for it. I had not thought of it before.

Finally, I think that juries are even less likely than judges to conclude that credentialed witnesses with scientific expertise are mistaken. The eye-opening article in this Symposium by criminal defense lawyer James Shellow claims, on the basis of experience, that jurors cannot understand or follow cross-examination aimed at revealing “flaws in methodology,” and that effective cross requires essentially peripheral tactics, such as attacks on character or a demonstration that the expert’s opinion is contradicted by published texts. The burden of the examples cited by Mr. Shellow is that the cross-examiner should exploit any unwillingness of the witness to acknowledge the authority of texts by casting that very fact as a demonstration of mendacity.¹⁴ This practitioner’s view supports empirical evidence described by Professor Sanders indicating that jurors do not effectively come to grips with scientific evidence.

II. *DAUBERT* DOES NOT THROW OUT TOO MUCH EVIDENCE

Daubert has been cast as one of the villains in toxic tort claims that fail,¹⁵ but I doubt that this claim is correct, and doubt even more that this claim shows that *Daubert* is in some serious sense misguided or mistaken. To start with, it was not clear on the day *Daubert* was decided whether the effect of the new doctrine was actually to raise or to lower the bar with respect to science (and now all expert testimony). Although a recent study concludes that *Daubert* subjects evidence proffered as science to increased scrutiny, *Daubert* itself threw out a ruling that excluded expert testimony, and some modern state decisions continue to declare that *Daubert* favors admissibility more than *Frye*.¹⁶

¹⁴ James M. Shellow, *The Limits of Cross-Examination*, 34 SETON HALL L. REV. ____ (upcoming in Fall 2003).

¹⁵ See generally Margaret A. Berger, *Upsetting the Balance Between Adverse Interests: The Impact of the Supreme Court’s Trilogy on Expert Testimony in Toxic Tort Litigation*, 64 LAW & CONTEMP. PROBS. 289, 318 (2001) (pointing out that *Daubert* led to exclusion of epidemiological testimony, criticizing opinions requiring that such proof show a doubling of relative risk, and suggesting that federal courts should look to state requirements for proving causation); Lucinda Finley, *Guarding the Gate to the Courthouse: How Trial Judges are Using their Evidentiary Screening Role to Remake Tort Causation Rules*, 49 DEPAUL L. REV. 335, 375 (1999) (urging that courts should not place the burden of scientific uncertainty on plaintiffs).

¹⁶ See *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 588 (1993) (contrasting the “liberal thrust” of the Rules and the “permissive backdrop” behind FRE 702 with the “austere standard” of *Frye*); *State v. Leep*, 569 S.E.2d 133, 143 (W. Va. 2002) (contrasting *Frye* with the “more liberal” *Daubert* standard and adopting the latter). But of course Judge Kozinski again excluded the evidence proffered in *Daubert*, this time applying the new standard. See *United States v. Daubert*, 43 F.3d 1311 (9th Cir. 1995); see also Lloyd Dixon & Brian Gill, *Changes In the Standards for Admitting Expert*

A. *Critical Look Approach*

Properly understood, *Daubert* only asks courts to look critically at evidence proffered as science (or expertise more generally), and to determine whether it is valid and what it can prove. In modern decisions *Daubert* does not automatically block efforts to prove causation by expert testimony resting, for example, on such techniques as animal studies¹⁷ or differential diagnosis.¹⁸ The former brings questions of dosage or exposure and questions stemming from differences between humans and animals. The latter involves attempts by treating physicians to eliminate other causes until only one explanation is left. Proof of this sort can survive scrutiny under *Daubert*, although it may properly be excluded if it fails adequately to fit the case, its factual basis is inadequate, or the methods or laboratory protocols were not properly followed. The problems of rational inference raised by such evidence virtually invite attempts to prove cause by evidence that does not really do so, and one cannot seriously argue that *all* such proof is a reliable indicator of cause. There is room for difference of opinion in applying the *Daubert* standard, and no doubt room for mistakes. But in areas of such difficulty, why would anyone expect otherwise?

It is true that proof based on chemical structure analysis has had tough sledding, but skepticism is justified by the fact that the technique is not suited to this use, and serves better as a tool for mapping out future research.¹⁹ It is also true that cases applying

Evidence In Federal Civil Cases since the Daubert Decision, 8 PSYCH. PUB. POL. & L. 251, 274 (2002) (commenting, *inter alia*, that standards for reliability “tightened in the years after the *Daubert* decision”).

¹⁷ *Compare* *Metabolife Int’l, Inc. v. Wornick*, 264 F.3d 832, 842 (9th Cir. 2002) (reversing summary judgment for defendants, in so-called “slap suit,” because trial judge erred in refusing to consider Asian animal studies in support of manufacturer’s claim that diet supplement was safe; *Daubert* “recognized that animal studies are not per se inadmissible and should be subjected to substantive analysis, just like other scientific evidence”), *and* *Curtis v M&S Petroleum, Inc.*, 174 F.3d 661, 669-70 (5th Cir. 1999) (partially reversing judgment dismissing claims arising out of exposure to benzene because testimony by plaintiff’s expert resting partly on animal studies satisfied *Daubert* standard), *with* *Rider v. Sandoz Pharm. Corp.*, 295 F.3d 1194, 1201 (11th Cir. 2002) (excluding testimony based on animal studies indicating that bromocriptine demonstrated vasoconstrictive properties in dogs and other animals because they did not suffice to indicate similar effects in humans).

¹⁸ *See* *Mattis v. Carlon Elec. Prod.*, 295 F.3d 856, 861 (8th Cir. 2002) (finding that testimony based on differential diagnosis can satisfy the *Daubert* standard).

¹⁹ *See* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 203 (Fed. Judicial Ctr. 1994) (stating that the Environmental Protection Agency uses structure activity relationships [“SARs”] in predicting toxicity of new chemicals, but “[their] reliability has a number of limitations”) [hereinafter REFERENCE MANUAL]; *see also* DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT

Daubert evince a preference for proving cause by means of epidemiological studies.²⁰ The problem with too much enthusiasm for epidemiological evidence is that it is often unavailable: Claimants cannot come up with such proof because it is expensive and takes a long time to develop. Even if such evidence constitutes “the gold standard” in this setting, however, courts applying *Daubert* regularly allow causation to be proved in other ways.

B. *Daubert as Source of Bad Rules*

Some commentators suggest that *Daubert* is the cause of certain “rules” that block recovery in toxic tort cases. One is the supposed rule that epidemiological evidence is admissible only if it shows a doubling of incremental risk. The problem is as follows: Suppose a study of two groups of 500 people, one group exposed to agent X and one not exposed. In the exposed group, we find that thirty-six suffer ailment Y, but in the unexposed group only twenty people suffer ailment Y. The usual standard of statistical significance, in which $p = .05$, requires that we be able to say that pure chance would produce the observed result only one time in twenty ($p = .05$ refers to that low probability).²¹ Under this standard, our result is significant.²²

TESTIMONY, *Toxicology: The Use of Toxicology in the Safety Assessment of Chemicals*, § 34-2.4 (2002) (stating that “SAR has the pitfall of the exquisite sensitivity of certain biological processes to relatively minuscule changes in chemical structure”); David E. Bernstein, *The Admissibility of Scientific Evidence after Daubert v. Merrell Dow Pharmaceutical, Inc.*, 15 CARDOZO L. REV. 2139, 2178 (1994) (stating that “[c]hemical structure analysis is an example of a scientific technique that has valid scientific uses but is not properly used to prove causal association, much less individual causation”); *Blum v. Merrell Dow Pharm., Inc.*, 705 A.2d 1314, 1323 (Pa. Super Ct., 1997) (finding that epidemiological studies “are necessary to establish causation,” and that chemical structure analysis and in vitro testing can only “confirm the biological plausibility of a causal relationship” but “contribute nothing” on their own).

²⁰ *Rider v. Sandoz Pharm. Corp.*, 295 F.3d 1194, 1199 (11th Cir. 2002) (considering “difficult question” whether evidence of causation “in the absence of epidemiology” can satisfy *Daubert*; here the answer is no; anecdotal evidence in the form of case reports, challenge/rechallenge data, chemical analogies, and animal studies are insufficient to prove that Parlodel causes strokes).

²¹ The calculations underlying the examples in this paragraph, and the text accompanying notes 68-71, *infra*, are the ones required in comparing the means of independent samples where standard deviation is not known (two groups of 500 people, one exposed to agent X, the other not exposed). The examples assume that prior research indicates a causal link between agent X and disease Y, which is important because the analysis must assume either that (1) it is unknown whether agent X might actually *lessen* the risk of disease Y or (2) it is known that agent X might increase the risk of disease Y and there is *no* reason to think it *lessens* that risk. In the latter situation, the analysis employs a one-tailed test, and more results survive scrutiny. See RUSSELL T. HURLBURT, *COMPREHENDING BEHAVIORAL STATISTICS*, 238-74

Scientists would take it seriously as an indication that agent X causes ailment Y.²³ Actually, it is now recommended that researchers reporting statistically significant findings include the “confidence interval” with their report. The latter describes *the range of outcomes* that would be expected to occur by pure chance no more than five percent of the time. The *narrower* the interval, and the further up from critical value that the interval lies, the higher the quality of the reported result.²⁴

Still, if all we knew about the plaintiff was that he was exposed to agent X and suffers ailment Y, even our statistically significant result does not *by itself* indicate that agent X probably caused plaintiff’s ailment. We could say otherwise, however, if the result showed more than a doubling of incremental risk—let us say that 42 people in the exposed group suffer ailment Y, and only twenty people in the

(3d ed. 2003) (providing an account of the underlying calculations). Following Professor Neil Cohen’s suggestion to use confidence intervals, rather than mere point estimates, the calculations in this article present both. See generally Neil Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385 (1985). For a critique of this approach, see D.H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54 (1987). For a reply, see Neil Cohen, *Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye*, 73 CORNELL L. REV. 78 (1987).

²² In this example (36 exposed people suffer ailment Y, and 20 unexposed people), the result is statistically significant at $p = .05$. For this level of significance, the critical point value of t is 1.65, and the observed point value of t , in the comparison of the two samples, is 2.13, which falls in the critical range. The confidence interval for the comparison is .007-.057. The null hypothesis is that exposure has no bearing on the number of ailing people. The null hypothesis assumes an observed value of t below 1.65, and assumes that the confidence interval will span the number 0 or fall below it. Since the observed point value of t exceeds the critical value of t , and since the confidence interval spans a range above 0, the result is statistically significant.

²³ I recognize that this standard has itself become controversial if taken as a minimum requirement for evidence offered in civil cases, and I return to this subject in Part III. See *infra* notes 68-80 and accompanying text.

²⁴ The reason a *narrow* confidence interval is better is that it indicates greater precision in the test. If results would be expected to exceed or fall below 14-15 only 5% of the time, the test is more precise than one in which the results would be expected to exceed or fall below 10-20 only 5% of the time. See HURLBURT, *supra* note 21, at 263 (noting with approval that journal editors “often require authors to report confidence intervals”). In the case of comparative risk, an interval spanning 1 (point of no correlation between exposure and ailment) would not be significant. See REFERENCE MANUAL, *supra* note 19, at 173 (explaining that with p value of .05, “a confidence interval would indicate the range of relative risk values that would result 95% of the time if the study were repeated,” so, the width of the confidence interval indicates “the precision of the point estimate of relative risk,” and narrower confidence intervals thus indicate more confidence in the resulting estimate; however, where interval spans critical value, a relative risk of 1.0, the results are not significant).

unexposed group. In this case, combining the epidemiological study with our knowledge of the plaintiff would suggest that agent X probably did cause his ailment. The reason is that more than half the observed instances of ailment Y in the exposed population were caused by agent X, so any one person in the group is more likely than not to have become ill from exposure to agent X. The result of this test, by the way, would once again satisfy the conventional notion of statistical significance.²⁵

In fact, however, the situation is seldom so simple. The probative force of such proof turns on such things as levels of exposure (time and dose), adequate sampling techniques, specificity of symptoms measured, and controlling for extrinsic (or potentially confounding) variables.²⁶ Moreover, taking seriously the conventions of statistical significance described above, even outcomes showing more than a doubling of incremental risk would not necessarily persuade scientists to draw any conclusions.²⁷ Also, a relative risk in the neighborhood of two is not as high as it sounds (scientists often see far higher relative risks).²⁸

²⁵ In this example (42 exposed people suffer ailment Y and 20 unexposed people), the result is statistically significant at $p = .05$. For this level of significance, the critical point value of t is 1.65, and the observed point value of t , in the comparison of the two samples, is 2.93, which falls in the critical range. The confidence interval for the comparison is .019-.069. Again the null hypothesis is that exposure has no bearing on the number of ailing people. The null hypothesis assumes an observed value of t below 1.65, and assumes that the confidence interval would span the number 0, or fall below it. Since the observed point value of t exceeds the critical value of t and since the confidence interval spans a range above 0, the result is statistically significant.

²⁶ See generally LEON GORDIS, EPIDEMIOLOGY 192-95, 204-17 (2d ed. 2000) (describing guidelines for studies of problems in causation, which require researchers to consider: (1) temporal relationship; (2) strength of association; (3) dose-response relationship; (4) replication of findings; (5) biologic plausibility; (6) consideration of alternative explanations; (7) cessation of exposure; (8) specificity of association; and (9) consistency with other knowledge). Other problems in epidemiological studies include: (a) selection bias; (b) information bias; (c) confounding factors; and (d) interaction. *Id.*; see also *Barrow v. Bristol-Myers Squibb Co.*, No. 96-689-Civ-Orl-19B, 1998 WL 812318, at *23 (M.D. Fla. Oct. 29, 1998). This case describes the Bradford-Hill criteria for appraising epidemiological proof of causation, which include:

- (1) the strength of the association or how far above 1.0 is the relative risk;
- (2) the consistency of the association or its reproducibility;
- (3) the specificity of the signs and symptoms or whether they are unusual and distinctive;
- (4) the dose response;
- (5) the temporality;
- and (6) the biologic plausibility of the theory of causation.

Id.

²⁷ See *infra* example 4, note 71 and accompanying text.

²⁸ See, e.g., *Hollander v. Sandoz Pharm. Corp.*, 289 F.3d 1193, 1212 (10th Cir. 2002) (noting study results that indicate that the relative risk of stroke in post-partum

Putting aside these problems, commentators are right that epidemiological studies should be admissible to show general causation even if relative risk is less than two, and a few decisions do miss this point.²⁹ When the purpose is to prove specific cause, it makes sense to insist on relative risk exceeding two because such proof is mathematically sufficient to satisfy the preponderance standard. But even this restriction assumes that there is no *other* evidence of exposure: If there is other evidence, then even epidemiological proof that does not show a doubling of risk is still relevant as *partial* proof of specific cause. Many modern decisions (state and federal) approve epidemiological evidence showing a relative risk less than two,³⁰ and many demonstrate good understanding of this idea.³¹ Decisions refusing to accept the proof when it shows a relative risk less than two often do so for other reasons, and not on the basis of simple misunderstanding.³²

women is 28.3, as compared with non-pregnant women); *Falise v. Am. Tobacco Co.*, 94 F. Supp. 2d 316, 336 (E.D. N.Y. 2000) (noting in cigarette smoking case that risks of lung cancer are 5 times higher for asbestos workers than for other workers, 50 times higher if asbestos workers smoke cigarettes, and 87 times higher if asbestos workers smoke more than a pack a day); *In re Joint Eastern and S. Dist. Asbestos Litig.*, 827 F. Supp. 1014, 1038 (S.D.N.Y. 1993) (reporting relative risk of lung cancer in cigarette smokers as compared to nonsmokers is "on the order of 10:1").

²⁹ *Sanderson v. Int'l Flavors & Fragrances*, 950 F. Supp. 981, 1000 (C.D. Cal. 1996) (holding that plaintiff's proof is not founded on epidemiological evidence showing relative risk greater than 2.0, "or some other evidence" of causation, so the evidence does not have "a valid scientific connection to the pertinent inquiry" under *Daubert* and stating that a relative risk of less than 2.0 "may suggest teratogenicity, but actually tends to disprove legal causation").

³⁰ Among federal cases, see *In re Hanford Nuclear Reservation Litig.*, 292 F.3d 1124, 1133 (9th Cir. 2002), which held that the trial court erred in requiring that epidemiological evidence show a relative risk greater than 2.0 and further stated that, to show "generic causation," plaintiffs only needed scientific evidence that radiation "was capable of causing" injuries such as those suffered by plaintiff. Among state cases, see *McDaniel v. CSX Transportation, Inc.*, 955 S.W.2d 257, 265 (Tenn. 1997), which adopted the *Daubert* standard, and also rejected the defense claim that epidemiological evidence is admissible only if it shows relative risk exceeding 2.0.

³¹ See *In re TMI Litig.*, 193 F.3d 613, 712 n.166 (3d Cir. 1999) (quoting passage from REFERENCE MANUAL); *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1315 n.16 (11th Cir. 1999) (finding a relative risk exceeding 2.0 permits "an inference that the plaintiff's disease was more likely than not caused by the agent," yet noting that no one doubts that smoking can cause heart disease even though relative risk in that setting is only 1.5); *In re Joint E. and S. Dist. Asbestos Litig.*, 827 F. Supp. 1014, 1028 (S.D.N.Y. 1993) (holding that an epidemiologist might find cause where relative risk is less than 2.0, but a "more likely than not" test "is not satisfied by epidemiological evidence alone" unless relative risk exceeds 2.0).

³² See *Allison*, 184 F.3d at 1315 (approving exclusion of evidence showing relative risk of 1.24 because "it was so significantly close to 1.0 that the court thought the study was not worth serious consideration for proving causation" in a breast implant case).

Another supposed rule holds that evidence of differential diagnosis cannot be admitted to show that agent X caused plaintiff to suffer ailment Y unless there is additional proof of general cause—whether epidemiological or based on animal studies—that agent X does cause ailment Y in some people. Such proof is popular because it is less expensive, and it can be provided by treating physicians who are not toxicologists or epidemiologists. Essentially the physician testifies that she tried to account for the ailment in other ways, through testing or treatment regimens, and thus eliminated all other possible or likely causes except agent X. Here the cases conflict. Some hold that differential diagnosis can only eliminate other causes, and because the technique is necessarily uncertain (it is hard to know when one has eliminated *all* but one cause), it can only supplement affirmative proof that agent X sometimes causes such ailments.³³ Other cases admit such proof without such preconditions, accepting it as relevant to show causation.³⁴ There is no settled rule, and courts seem to be trying hard to distinguish between testimony that does eliminate other plausible risks, and testimony that does not.³⁵ In other words, the decisions seem to draw sensible qualitative distinctions that connect with the task that *Daubert* asks courts to perform.³⁶

³³ See e.g., *Rider*, 295 F.3d at 1194, 1199 (affirming summary judgment for defendant in a Parlodel suit after a *Daubert* hearing because the proffered evidence of causation was not sufficient to prove causation; also finding that case reports based on differential diagnosis could not by themselves prove causal link “because they report symptoms observed in a single patient in an uncontrolled context,” which can “rule out other potential causes” but cannot rule out the possibility that the observed effect “is simply idiosyncratic or the result of unknown confounding factors,” so such reports “may support other proof of causation,” but “ordinarily cannot prove causation” by themselves); *Hollander*, 289 F.3d at 1210-11; *Glastetter v. Novartis Pharm. Corp.*, 252 F.3d 986, 988 (8th Cir. 2001) (containing an analysis similar to *Rider*).

³⁴ *Mattis v. Carlton Elec. Prods.* 295 F.3d 856, 861 (8th Cir. 2002) (admitting differential diagnosis testimony in electrician’s suit against a maker of PVCs as sufficient to prove causation, where a physician “ruled out other possible causes,” including “smoking, asthma, or ammonia,” and concluded that plaintiff developed reactive airways syndrome “as a result of his exposure to Carlon cement fumes”).

³⁵ See Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis Testimony to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law*, 64 LAW & CONTEMP. PROBS. 107, 137 (2001) (stating that *Daubert* has led to “greater skepticism” about differential diagnosis testimony, but that courts have reached a “fair degree of consensus” on questions such as whether the proponent must first offer “ruling-in” evidence before offering “ruling-out” testimony and on the sufficiency of proof resting largely on “temporal order”).

³⁶ *Cooper v. Smith & Nephew, Inc.*, 259 F.3d 194, 202 (4th Cir. 2001) (stating that differential diagnosis “normally should not be excluded because the expert has failed to rule out every possible alternative cause,” but that testimony may be

Of course both of these “rules” could be deployed unwisely to block just claims. It might even be true that a *Daubert* regime provides greater opportunity for courts to make such mistakes than *Frye* did. Certainly the most liberal “Rule 702” standard in current use, in which courts look at credentials and apply the “helpfulness” standard, would less often lead to exclusion of such evidence. But such proof should be excluded when it is thin, and looking directly at the science seems a good thing, not a bad thing.

These supposed “rules” also invite the criticism that courts are applying *Daubert* to measure the sufficiency of scientific evidence, rather than its relevancy.³⁷ Some opinions appear to collapse notions of relevancy and sufficiency. That is not the fault of *Daubert*, however, and other modern decisions clearly understand the difference between relevance and sufficiency in this setting.³⁸ In partial answer to this criticism, it is worth noting that a court asked to rule on an offer of proof sometimes *should* exclude the evidence because it is insufficient. In other contexts, it is perfectly conventional for courts to sustain objections on the ground that evidence does not suffice to prove the point for which it is offered.³⁹ There is absolutely nothing wrong with doing so, at least in cases in which the proponent has no additional evidence on the point in question and has had an adequate opportunity to advise the court about the proof that he does have and to make the usual proffer.

excluded where it fails to consider other potential causes). The court in *Cooper* found that the record was “replete with evidence that smoking can cause non-unions to occur,” and that plaintiff was a pack-a-day smoker for 25 years, a fact that expert “categorically dismissed.”

³⁷ See generally Edward J. Imwinkelried, *Daubert Revisited: Disturbing Implications*, 22 CHAMPION 18 (May 1998) (criticizing federal courts that have invoked the “fit” or “relevancy” prong of *Daubert* to insist that epidemiological proof satisfy what amounts to a sufficiency standard).

³⁸ See Joint Dist. Asbestos Liab. Litig., 52 F.3d 1124, 1133, 1137 (2d Cir. 1995) (stating that “*Daubert* did not alter the traditional sufficiency standard” and reversing the trial court’s judgment for defendant as matter of law because the trial court “erred in ruling that plaintiff presented insufficient epidemiological and clinical evidence” to prove causation).

³⁹ See, e.g., *Tennison v. Circus Circus Enters., Inc.*, 244 F.3d 684, 690 (9th Cir. 2001); *Viking Theatre Corp. v. Paramount Film Distrib. Corp.*, 320 F.2d 285, 296 (3d Cir. 1963), *aff’d*, 378 U.S. 123 (both rejecting offers of proof because the evidence could not prove the point for which it was offered unless other evidence was offered, which offeror did not include or could not obtain).

C. Daubert and Erie

Professor Margaret Berger offers another criticism of what she takes to be the judge-made doubling rule.⁴⁰ Imposing this rule in federal diversity suits, she argues, violates the *Erie* doctrine. Stated in its strongest terms, the argument is that this judge-made rule, created in a *Daubert*-inspired construction of FRE 702, is substantive and is intended to affect outcome in a particular class of cases. Drawing on a modern opinion by Judge Posner in the *Healy* case, Professor Berger suggests that federal courts must apply any state substantive rule that is “in actual conflict” with a Federal Rule, and any “state procedural rule” that applies to “a particular substantive area.”⁴¹ Lest anyone think the decision in *Hanna* stands in the way because it puts the Federal Rules beyond *Erie*-based challenge, Professor Berger reminds us that a prominent modern decision requires federal courts to apply state substantive law even when a Federal Rule is in play.⁴² She concludes that when a state court interprets evidence law “so as to better a plaintiff’s odds of prevailing in toxic tort litigation,”⁴³ the result is a state rule applying in a particular class of cases, so federal courts must observe it. Perhaps more importantly, the state rule is substantive because it creates an incentive for manufacturers “to take more care in testing their products.”⁴⁴

This is a brave and inventive argument. There is something to be said for the proposition that if the state and federal systems persistently produce different outcomes in similar cases, on account of what seems to be different standards of proof, while purporting to apply the same substantive principles, the result would be troublesome. I concur in Professor Berger’s argument that federal courts should consider state precedents on matters closely related to sufficiency standards, and on evidential conventions that seem closely

⁴⁰ See generally Berger, *supra* note 15.

⁴¹ See *S.A. Healy Co. v. Milwaukee Metro. Sewerage Dist.*, 60 F.3d 305, 309 (7th Cir. 1995), cited twice with approval in *Gasperini v. Center for Humanities*, 518 U.S. 415, 428 n.7, 429 (1996). *Erie* refers to *Erie R.R. Co. v. Tompkins*, 304 U.S. 64 (1938).

⁴² The reference in this text is to *Hanna v. Plumer*, 380 U.S. 460 (1965). The modern decision is *Gasperini v. Ctr. for Humanities*, 518 U.S. 415, 419 (1996) (holding that, in diversity suits, federal courts must apply state statute controlling compensation awards for excessiveness or inadequacy; noting, however, that the statute directs appellate courts to exercise this power, while in the federal system, the trial judge must take this responsibility).

⁴³ Berger, *supra* note 15, at 319.

⁴⁴ The indicated conclusion is that a contrary judge-made federal rule violates the Rules of Decision Act, 28 U.S.C. § 1652 (2003), or the Rules Enabling Act, 28 U.S.C. § 2072(b) (2003) in its modern formulation. Berger, *supra* note 15, at 312-19.

related to substantive principles. Federal courts often do just that,⁴⁵ but most of the time they do not view themselves as bound by state law, and it seems telling that Judge Posner himself goes to great lengths to avoid being bound by state rules relating to proof in tort cases.⁴⁶

In the end, the *Erie* argument seems misconceived. To begin with, federal decisions don't impose a "doubling rule" for epidemiological evidence (instead they analyze and assess probative worth and sufficiency). More importantly, different results on this question seem epistemological, rather than policy-driven. What I mean by epistemological is that the differences look like variations in attempts by federal and state judges to implement the "sufficient evidence" standard by deciding "how much evidence is enough" to allow a reasonable juror to find that cause has been proved under the preponderance standard. If the results were policy-driven, one would expect to see opinions linking the discussion of "how much is enough" to particular substantive standards, or to the purposes of tort law as compensatory and loss-spreading or as shifting to manufacturers only actual costs of injury while keeping innovation alive and costs down. One would also expect policy-based decisions to generate bright-line rules, or even statutes governing recurrent situations.

Of course judges might mask their decisions, justifying them

⁴⁵ See *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1320 (11th Cir. 1999) (noting in a breast implant case that *Erie* requires application of state substantive standards and commenting, with reference to evidence of relative risk, that state law requires proof based on "reasonable medical probability," interpreted to mean "the functional equivalent of preponderance of the evidence"); *In re Simon II Litig.*, 211 F.R.D. 86, 157 (E.D. N.Y. 2002) (commenting that allowing "statistical proof" of causation in cigarette smoking litigation does not conflict with *Erie* because there is "no ruling New York case which holds that state substantive law will not permit the use of modern aggregation forensic tools to support a massive fraud action").

⁴⁶ See *S.A. Healy Co.*, 60 F.3d at 309-10. Judge Posner, writing the opinion in *Healy*, cites two tort cases in which the *Erie* issue is "pretty easy" because the state rule is "limited to a particular substantive area." *Id.* Judge Posner authored both prior opinions, and both times he *avoided* applying the state rule. See also *Barron v. Ford Motor Co. of Canada, Ltd.*, 965 F.2d 195 (7th Cir. 1992) (holding that a state statute blocking proof that claimant was not wearing seatbelt was substantive, but that it did not apply where defendant claimed that seatbelts were a design element relevant to the question whether it was reasonable to make sunroof out of laminated glass); *Flaminio v. Honda Motor Co.*, 733 F.2d 463, 471-72 (7th Cir. 1984) (refusing to apply a state rule letting plaintiff prove design change, because a federal rule blocks it; noting that the matter is both substantive and procedural, but that it would be "melodramatic" to label federal rule as substantive and require federal courts to apply state counterpart) (at the time, state and federal rule were textually identical, but state and federal courts diverged on question whether rule excludes such proof).

epistemologically rather than by reference to substantive policy. Indeed, humanly speaking it is hard to imagine deciding whether indirect proof of cause suffices *unless* one also thinks about substantive policies, and what is humanly at stake—a serious injury or ailment on one side, the future of a drug on the other. But if *every* judicial attempt to think epistemologically (hence procedurally) is viewed as a masked effort to implement policy choices, then the *Erie* doctrine is doomed. We would be forced to the conclusion that federal courts cannot at the same time operate as “an independent system for administering justice” as contemplated by Justice Brennan in *Byrd*⁴⁷ while complying with *Erie*’s command to honor state substantive policy choices. Professor Berger does not make such an extravagant claim, but her contention that federal judges are sometimes implementing substantive policy choices comes close to that, since the opinions do not *say* that they are behaving in this way.⁴⁸

In any event, the federal decisions that have noticed this issue have mostly avoided concluding that *Erie* mandates following state practice on this point.⁴⁹ On the root question whether a state or federal standard governs the sufficiency question in diversity litigation, there is stronger support for the proposition that federal law governs than there is for applying a state standard.⁵⁰

⁴⁷ *Byrd v. Blue Ridge Rural Elec. Coop., Inc.*, 356 U.S. 525, 537 (1958).

⁴⁸ See Berger, *supra* note 15, at 301-06 (noting that in federal decisions “insisting on epidemiological proof” or those insisting that a plaintiff’s epidemiological evidence show a relative risk exceeding 2.0, judges “are not making value-free determinations that are the inevitable consequences of a system of rational proof”).

⁴⁹ See *Bartley v. Euclid, Inc.*, 158 F.3d 261, 272-73 & n9 (5th Cir. 1998) (finding that if state law requires the evidence to show more than doubled risk in the exposed population, the proof satisfied the standard; if this requirement defines burden of proof, it is arguably “procedural rather than substantive, and therefore controlled by federal rather than state law”) (dissent argues that state rule is substantive); *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris, Inc.*, 178 F. Supp. 2d 198, 259 (rejecting defense claim, in smoking case, that *Erie* required federal courts to require “proof of individual injury,” and commenting that the question is “better posed as a question of legal sufficiency”; if mixture of statistical and individualized evidence can prove cause, then “no *Erie* question is presented by federal evidentiary procedures which allow for the use of aggregate proof”); *Hall v. Baxter Healthcare Corp.*, 947 F.Supp. 1387, 1394-95 (D. Or. 1996) (rejecting a claim that *Erie* required the federal court to apply state rule relating to proof of causation in breast implant suit). *But see Nat’l Bank of Commerce v. Assoc. Milk Producers, Inc.*, 22 F. Supp. 3d 942, 948 n.4 (E.D. Ark. 1998) (stating that if the application of a “federal evidentiary rule” leads to dismissal, whereas the application of the state rule would not, “then, under *Erie*, the evidentiary ruling might be considered substantive rather than procedural”); *Raynor v. Merrell Pharm.*, 104 F.3d 1371, 1376 (D.C. Cir. 1997) (suggesting that a “question of sufficiency would be a substantive rule under *Erie*”).

⁵⁰ See generally 9A CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2525 (2d ed. 1995) (opining that many courts now agree that

D. *The 2000 Amendments*

The amendments to FRE 702 adopted in 2000 reinforce *Daubert's* message that courts are to take a close and independent look at evidence proffered as science, and increase the difficulties of arguing under *Erie* that rulings on sufficiency are substantive. In effect, these amendments say that everything that could affect validity and accuracy count. Not only should courts insure that scientific testimony "is the product of reliable principles and methods," which was the actual holding of *Daubert*, and not only should courts insure that scientific evidence rests on "sufficient facts or data," which was part of the sufficiency calculus that courts perform under FRCP 50 and sometimes in connection with rulings on offers of proof under FRE 103, but courts should *also* take steps to ensure that the witness "has applied the principles and methods reliably to the facts."⁵¹

The latter provision resolves a conflict among the cases in favor of more judicial scrutiny. This language directs trial judges to consider issues of laboratory protocol in determining whether to admit or exclude expert testimony, meaning that these issues affect not merely weight, but admissibility. Some pre-amendment authority had pointed toward this conclusion, but other decisions pointed toward the opposite conclusion.⁵² Whether the 2000 amendments

"principle seems to require that the federal court apply the federal test") (internal citation omitted); *Daniels v. Twin Oaks Nursing Home*, 692 F.2d 1321, 1323-24 (11th Cir. 1983) (stating that it is settled under *Erie* that "federal law controls questions of the sufficiency of the evidence in state law claims"); *Boeing Co. v. Shipman*, 411 F.2d 365, 368-69 (5th Cir. 1969) (holding that "in diversity cases federal courts apply a federal rather than a state test for the sufficiency of evidence to create a jury question") (en banc). *Contra* *Burke v. Deere & Co.*, 6 F.3d 497, 511 (8th Cir. 1993) (holding that state law determines sufficiency).

⁵¹ FED. R. EVID. 702. The Advisory Committee Note accompanying the 2000 amendments to FRE 702 states that it is "important" that the "application" of principles and methods "be conducted reliably," and cites an opinion by Judge Becker. As the author of the opinion in *Downing*, which anticipated *Daubert* and was cited there with approval, Judge Becker has been unusually active and innovative in dealing constructively with problems of scientific evidence. *See* *United States v. Downing*, 753 F.2d 1224 (3d Cir. 1985). Elsewhere Judge Becker endorsed the proposition that judges should assess not only validity and accuracy of principles, but issues of application. *See* *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 745 (3d Cir. 1994) (finding that "any step" making expert testimony "unreliable" also makes it inadmissible, regardless whether it "completely changes a reliable methodology or merely misapplies" it).

⁵² *Compare* *United States v. Martinez*, 3 F.3d 1191, 1197-98 (8th Cir. 1993) (stating that under *Daubert* the court should require expert to show that he "properly performed the protocols involved in DNA profiling"), *with* *United States v. Chischilly*, 30 F.3d 1144, 1154 (9th Cir. 1994) (finding that questions relating to conduct of laboratory procedures go to weight, not admissibility), *and* *United States v. Shea*, 211 F.3d 658, 668 (1st Cir. 2000) (stating that flaws in application of methodology go to

will have real impact on the way courts deal with scientific evidence has yet to be seen. These changes have not yet been widely adopted by the states, perhaps because they have not had enough time to consider the matter.⁵³

III. *DAUBERT* PROPERLY CONCEIVES SCIENCE, AND TAKES THE RIGHT LEGAL STANCE TOWARD SCIENCE

Critics complain that *Daubert* is incoherent, perhaps even internally conflicted, in two critical respects—in its view of science, and in its conception of the proper relationship between science and law.⁵⁴ There is power in these observations, but I mean to say once again that *Daubert* is not at fault. Indeed, one of the strengths of the opinion is that its vision is broad enough to embrace internal tensions and difficulties in science, and in the relationship between law and science, that cannot be avoided. For us in the scholarly community, and for judges toiling in the vineyards, the “task at hand”⁵⁵ is to make our way toward appropriate accommodations of

weight, not admissibility); *also compare* *People v. Castro*, 545 N.Y.S.2d 985, 996 (N.Y. Sup. Ct. 1989) (holding that pretrial hearing should determine whether “the experiments and calculations performed by the testing laboratory in the particular case yielded results sufficiently reliable to be presented to the jury,” and factual issues relating to “the reliability of any particular test” can affect weight, but can also show that the evidence is “inadmissible as a matter of law”), *with* *Fishback v. People*, 851 P.2d 884, 893 (Colo. 1993) (matters of “implementation and execution” go to weight, not admissibility) (applying *Frye* standard). *See also* *People v. Shreck*, 22 P.3d 68, 73-9 (Colo. 2001) (discarding *Frye* and adopting standard similar to *Daubert*, noting that some courts consider that matters of implementation of methods affect admissibility, but not taking a position on this issue); Edward J. Imwinkelried, *The Debate in the DNA Cases Over the Foundation for the Admission of Scientific Evidence: The Importance of Human Error as a Cause of Forensic Misanalysis*, 69 WASH. U. L.Q. 19 (1991) (matters of laboratory protocol should affect admissibility, not just weight).

⁵³ I am aware of one state that has apparently adopted the new language. *See* Miss. R. EVID. 702 (adopting the new federal language). In Colorado, the state Supreme Court rejected a recommendation by its Evidence Rules Advisory Committee to adopt the federal language, after I argued unsuccessfully in favor of its adoption. The expressed concern was that adopting language that seemed so closely related to *Daubert* would essentially adopt *Daubert* itself, a questionable position given that Colorado had just rejected the *Frye* standard, in order to adopt its own standard, which is similar to, but not identical with, *Daubert*. *See* *Shreck*, 22 P.3d at 73-79 (requiring trial courts to consider reliability of expert testimony, qualifications of witness, and usefulness of testimony, thus endorsing the *Daubert* factors without adopting *Daubert*).

⁵⁴ *See* Jan Beyea and Daniel Berger, *Scientific Misconceptions Among Daubert Gatekeepers: The Need for Reform of Expert Review Procedures*, 64 LAW & CONTEMP. PROB. 327 (2001); *see also* Margaret Farrell, *Daubert v. Merrell Dow Pharmaceuticals, Inc.: Epistemology and Process*, 15 CARDOZO L. REV. 2183 (1994).

⁵⁵ *Daubert*, 509 U.S. at 597 (characterizing the responsibility of trial judge to include ensuring that an expert’s testimony “rests on a reliable foundation and is

these difficulties.

Let us consider first the charge that *Daubert's* view of science is incoherent. On the one hand, we can see in the core of the opinion (the part that adopts the validity standard and speaks of accuracy and propositions that can be tested or “falsified”) an apparent belief that science is a static body of objective knowledge reflecting certainty. On the other hand, we also find in *Daubert* suggestions that (a) science is a process, hence anything but static; (b) scientific knowledge does not reflect certainty, but is uncertain and contingent; and (c) scientific expertise is affected by the forces that generate litigation, hence *subjective* in some respects, and socially constructed.⁵⁶ This incoherent view, it is said, makes the task that *Daubert* sets for judges impossible to perform: In effect, *Daubert* charges them to apply static objective standards in appraising shifting subjective, contingent knowledge.

Let us consider the relationship between law and science, as *Daubert* envisions it. On the one hand, *Daubert* affirms that it is the job of courts to appraise science, and courts are not simply to defer to the scientific community on the question whether evidence presented as science is valid and reliable. This role for courts is what we mean by “gatekeeping.” On the other hand, *Daubert* says courts are to judge science by the standards that *scientists* deploy in judging science. *Kumho Tire* adds an exclamation point in commenting that scientists are to bring to the courtroom “the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.”⁵⁷ Again this incoherent view asks courts to do what they cannot do and fails to recognize that science and law have different agendas, goals and purposes, and operate under different constraints.

A. A Defective View of Science?

On the question whether *Daubert* has a defective view of science, I would begin by suggesting that the problem of objectivity has a familiar ring, perhaps because bridging the gap between human perceptions and the world has engaged philosophers for thousands of years, and the conversation is not over yet. How surprising is it to

relevant to the task at hand”); *see also Kumho Tire*, 526 U.S. at 152 (explaining how judges apply *Daubert* factors “to the case at hand”).

⁵⁶ *Daubert*, 509 U.S. at 590, 596 (noting that “arguably, there are no certainties in science,” and commenting that science is “a *process* for proposing and refining theoretical explanations about the world,” making it qualitatively different from the law).

⁵⁷ *Kumho Tire*, 526 U.S. at 152.

find that philosophers and historians of science report that science too is not the wholly objective edifice that we outsiders envision? Thomas Kuhn's salient work argues that the choice between what he called "competing paradigms" in science "cannot be determined merely by the evaluative procedures characteristic of normal science."⁵⁸ The philosopher of science Karl Popper, of whom we have heard because *Daubert* draws on his work, takes a similar position. In defending the proposition that science is distinguished from other forms of knowledge by the fact that it can be "falsified," he argued that a proposition can be falsified only by a "basic statement," meaning a singular empirical statement that is accepted because it has been tested "inter-subjectively" rather than objectively.⁵⁹

Yet views as skeptical as these cannot long survive unchallenged in a world that has seen such extraordinary accomplishments as laser surgery, the internet, space stations and jet airline travel. Obviously science has answers to critical questions, and Kuhn and Popper both recognized as much. What Kuhn called "normal science" proceeds, he wrote, out of random early "fact-gathering" toward something resembling "an accepted model or pattern" that he called a "paradigm," which succeeds because it is more successful than other paradigms "in solving a few problems that the group of practitioners has come to recognize as acute." "Mopping-up" operations that are "what engage most scientists throughout their careers," and these proceed after the adoption of a paradigm, and constitute "normal science."⁶⁰ In the end, what counts most as a critical criterion of scientific paradigms is predictive accuracy.⁶¹ Popper was less direct in suggesting a positive account of scientific knowledge, but he did comment that scientists reach a kind of stopping point with "statements about whose acceptance or rejection the various investigators are likely to reach agreement," and he acknowledged that we must find such stopping points or end in a new "Babel of

⁵⁸ THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* 94 (3d ed. 1996).

⁵⁹ KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* §§ 8, 22, 28, 29 (Routledge Classics 2002) (stating that a theory can be "falsified" only by means of "a *reproducible effect* which refutes the theory").

⁶⁰ KUHN, *supra* note 58, at 15, 23, 25-27 (these "mopping-up" operations involve investigating those facts that "the paradigm has shown to be particularly revealing of the nature of things," as well as facts that lack "intrinsic interest" but "can be compared directly with predictions" from the paradigm).

⁶¹ THOMAS S. KUHN, *Objectivity, Value Judgment, and Theory Choice*, in *INTRODUCTORY READINGS TO THE PHILOSOPHY OF SCIENCE* 436 (Klemke et al. eds., 3d ed. 1998) (listing "as characteristics of a good scientific theory" the following factors: accuracy, consistency, breadth of scope, and fruitfulness in the sense of encouraging new phenomena or previously unnoticed relationships).

Tongues” in which scientific discovery “would be reduced to absurdity,” and “the soaring edifice of science would soon lie in ruins.”⁶²

Perhaps *Daubert* should be faulted for being too skeptical of science. As Professor Ron Allen argues, it seems odd to posit as a standard of scientific validity the question whether a proposition can be “falsified” as opposed to “verified” or “confirmed.” Arguably such a standard is too demanding if we wind up accepting scientific knowledge only if it has been tested in every conceivable way,⁶³ and Thomas Kuhn took issue with the very idea of “falsification” in an account that stresses more positive notions of verification.⁶⁴ But I do not believe *Daubert* meant to erect a barrier as high as that. It may be comforting to consider that Popper, in explaining why he chose “falsifiability” as the central criterion of science, says his purpose was to distinguish science from myth and metaphysics. Specifically he had in mind the claims of Carl Marx, Sigmund Freud, and Alfred Adler, which he viewed as “pseudo-science” more closely resembling astrology than astronomy.⁶⁵ Popper did not actually argue that nothing could be accepted until it was tested so exhaustively that nobody could doubt it. Rather, he argued that a scientific

⁶² POPPER, *supra* note 59, § 29 (characterizing this situation as “a failure of language as a means of universal communication”).

⁶³ See Ronald J. Allen, *Expertise and the Daubert Decision*, 84 J. CRIM. L. & CRIMINOLOGY 1157, 1169-71 (1994) (arguing that *Daubert* “adopted uncritically the view that Popperian falsifiability is at the heart of modern science,” showing “no awareness” that that view is controversial and inadequate because it suggests that science produces knowledge only if it survives “all conceivable tests,” but fails to account for the accomplishments of scientists, who “do not believe that all they know are negatives,” and know “a lot of positive truths” too). For a reply to this criticism, see Sean O’Connor, *The Supreme Court’s Philosophy of Science: Will the Real Karl Popper Please Stand Up?*, 35 JURIMETRICS J. 263 (1995).

⁶⁴ KUHN, *supra* note 58, at 145-47 (describing “probabilistic verification theories” that “compare the given scientific theory with all others that might be imagined to fit” the data, or construct by imagination “all the tests that the given scientific theory ‘might conceivably be asked to pass,’ and doubting that any test can falsify any theory because “no theory ever solves all the puzzles,” and indeed “it is just the incompleteness and imperfection of the existing data-theory fit that, at any time, define many of the puzzles that characterize normal science”).

⁶⁵ Karl Popper, *Science: Conjectures and Refutations*, in INTRODUCTORY READINGS TO THE PHILOSOPHY OF SCIENCE, 38-40 (Klemke et al. eds.) (describing a conversation with Adler in which the author had mentioned a child, whom Adler had “no difficulty in analyzing in terms of his theory of inferiority feelings, although he had not even seen the child,” because, as Adler said, he had “thousandfold experience,” leading the author to conclude that that Adler’s “previous observations may not have been much sounder than this new one,” proving only that any case “could be interpreted in light of the theory”).

proposition is one that *can be* tested.⁶⁶ And of course the Popper account tracks a salient feature of the scientific method, which is to test a hypothesis to see whether experimental results refute it (testing the “null hypothesis” in common parlance).

In sum, what we might take as incoherence or internal conflict in *Daubert’s* view of science can also be understood more constructively as a kind of dualism that embodies a view of science similar to what we find in Kuhn and Popper. Rather than abandoning any search for a validity standard, this dualistic view should lead us to recognize, in words that Professor Nance might find congenial,⁶⁷ that reliability is not an all-or-nothing concept, but a relative concept: Often it will be possible to insist on a kind of “certainty” of the sort that we have in mind when we speak of the tides or the hour of sunrise, but other times we can only expect the sort of “certainty” that we have when we say that asbestos causes some kinds of lung cancer.

B. Misconceived Relationship between Law and Science?

A serious criticism of *Daubert* is that courts are being led to demand a higher level of statistical significance than is appropriate. Epidemiological evidence might support the conclusion that exposure to agent X increases the risk of ailment Y, thus in turn supporting an inference of general causation (agent X causes some instances of ailment Y). Alternatively, it might support the conclusion that exposure more than doubles the risk, thus in turn supporting an inference of general causation *and* perhaps even specific causation (plaintiff was exposed and is ailing, and so agent X is the cause).

To illustrate these points, consider some examples comparing 500-person samples (one group exposed to agent X, one not exposed). Example 1 (risk increase): We find that thirty-six exposed people have ailment Y, but only twenty unexposed people. These numbers suggest that ailment Y suffered by sixteen out of thirty-six

⁶⁶ LOGIC OF SCIENTIFIC DISCOVERY, *supra* note 59, § 22 (we must “clearly distinguish between falsifiability and falsification”); *see also* *Conjectures and Refutations*, *supra* note 65, at 43 (task is not to identify “meaningfulness or significance” or “truth” or “acceptability,” but rather to distinguish statements and systems belonging to “the empirical sciences” from all others, whether “psychoanalytic” or “myth” or something else; the latter are *not* “unimportant, or insignificant,” and “may contain important anticipations of scientific theories”; indeed, psychoanalytic theories “contain most interesting psychological suggestions, but not in a testable form”).

⁶⁷ *See* Dale A. Nance, *Reliability and the Admissibility of Experts*, 34 SETON HALL L. REV. __ (upcoming in Fall 2003).

exposed people came from agent X, that relative risk is 1.8, and that there is a .444 probability that agent X caused any given case of ailment Y in the exposed population.⁶⁸ Example 2 (risk increase): In the exposed group, twenty-nine people have ailment Y; again only twenty in the unexposed group. These numbers suggest that ailment Y suffered by nine out of twenty-nine exposed people came from agent X, that relative risk is 1.45, and that there is a .310 probability that agent X caused any given case of ailment Y in the exposed population.⁶⁹ Example 3 (risk more than doubled): We find that forty-two exposed people have ailment Y, but only twenty unexposed people have it. These numbers suggest that the ailments suffered by twenty-two out of forty-two exposed people came from agent X, that relative risk is 2.1, and that there is a .542 probability that agent X caused any given case of ailment Y in the exposed population.⁷⁰ Example 4 (risk more than doubled): We find that five exposed people have ailment Y, and only two unexposed people. These numbers suggest that the ailments suffered by five out of seven exposed people came from agent X, that relative risk is 3.5, and that there is a probability of .714 that agent X caused any given case of ailment Y in the exposed population.⁷¹

⁶⁸ In this example (36 exposed and 20 unexposed people suffer the ailment) recall from note 22 that the result is statistically significant at $p = .05$. The observed value of t exceeds the critical value of t , and the confidence interval for the comparison is .007 to .057, which lies above the value of 0 that the null hypothesis would assume.

⁶⁹ In this example (29 exposed and 20 unexposed people suffer the ailment) the result is *not* statistically significant at $p = .05$. For this level of significance, the critical point value of t is 1.65, and the observed point value of t , in the comparison of the two samples is 1.28, which falls below the critical range. The confidence interval for the comparison of the two samples is -.005-.041. Again the null hypothesis is that exposure has no bearing on the number of ailing people. The null hypothesis assumes an observed value of t below 1.65, and assumes that the confidence interval will span the number 0. Since the observed point value of t is less than 1.65, and since the confidence interval *does* span 0, the result is *not* statistically significant (we do not reject the null hypothesis).

⁷⁰ In this example (42 exposed people suffer the ailment, and 20 unexposed people), recall from note 25 that the result is statistically significant at $p = .05$. The observed value of t exceeds the critical value of t , and the confidence interval for the comparison is .019-.069, which lies above the value of 0 that the null hypothesis would assume.

⁷¹ In this example (5 exposed people suffer the ailment, and 2 unexposed people), the result is *not* statistically significant at $p = .05$. For this level of significance, the critical point value of t is 1.65, and the observed point value of t , in the comparison of the two samples, is 1.2, which falls below the critical range. The confidence interval for the comparison of the two samples is -.002-.014. Again the null hypothesis is that exposure has no bearing on the number of ailing people. The null hypothesis assumes an observed value of t below 1.65, and assumes that the confidence interval will span the number 0. Since the observed point value of t is less

The usual approach to the question whether agent X causes ailment Y is to begin with the “null hypothesis” that there is no correlation. If that were so, then in the two samples (500 exposed people; 500 unexposed people) the number of observed instances of ailment Y would be the same. But the outcome of studies may suggest, as do Examples 1-4 above, that the null hypothesis should be rejected, and that agent X does indeed cause ailment Y. Let us begin by understanding the meaning of these outcomes. To start with, the suggested conclusions are all general and qualified, and each can be deployed to state a probability, but not a certainty. The conclusions are all general because they suggest that agent X is a causal factor in the mass of observed instances of ailment Y. The conclusions are qualified because they suggest that agent X causes certain percentages of observed instances of ailment Y in exposed populations. The conclusions can be deployed to state probabilities because they suggest that, among ailing and exposed people, there are certain probabilities, equal to the percentages suggested by the figures, that any given instance of ailment Y was caused by agent X.

In the description set forth above, Examples 1 and 2 indicate that exposure raises the risk but does not double it, thus supporting inferences of general cause. Examples 3 and 4 indicate that exposure more than doubles the risk, thus supporting inferences of both general and specific cause. Epidemiologists would likely *accept* the conclusions indicated by Examples 1 and 3, because they are statistically significant at the level of $p = .05$, meaning that there is but one chance in twenty that the numbers would appear by chance—by random and inevitable differences in 500-person samples taken from the general population. Epidemiologists would likely *reject* the conclusions indicated by Examples 2 and 4 because they are *not* statistically significant at this level. The effect of this convention is to discourage “false positives” in favor of “false negatives,” or (as it is sometimes said) to discourage α -errors by incurring more β -errors, or discourage Type I errors in favor of more Type II errors. In other words, the conventional standard is less tolerant of errors that would *find* a causal connection and more tolerant of errors that would *fail to find* one. Examples 2 and 4, which suggest respectively that twenty-nine exposed people have ailment Y (as opposed to twenty unexposed people), and that five exposed people have ailment Y (as opposed to two unexposed people), would not be accepted because the results do not satisfy the conventional standard of statistical

than 1.65, and since the confidence interval *does* span 0, the result is *not* statistically significant (we do not reject the null hypothesis).

significance.

It is important to note that the conventional standard does *not* mean that science usually accepts only results that are ninety-five percent certain. Statistical significance at the level of $p = .05$ means that there is but one chance in twenty that the observed results could happen by chance. *Satisfying* the standard means that there is one chance in twenty (or less) that mere accidental variation would produce such a result, *not* that we can be 95% certain that the outcome (twenty-two out of forty-two observed ailments were caused by exposure, or sixteen out of thirty-six, as Examples 1 and 3 indicate) is correct. We do not and cannot know that. All we know is that the observed outcome would rarely be produced by chance alone, which gives us *some reason* to believe that the indicated correlation is correct. Any suggestion, however, that the conventional standard produces results of which we are ninety-five percent certain is false.⁷²

Hence, it still needs to be said, it would make *no* sense to suggest that the civil justice system should accept results that are statistically significant at, say, the level $p = .40$ since such results leave us sixty percent certain of the conclusion, thus easily satisfying the notion of a preponderance of the evidence. Suggestions of this sort are close akin to the “prosecutor’s fallacy,” to which courts and lawyers sometimes fall prey. That fallacy equates the inverse of a scarcity factor with the probability of guilt: “The evidence shows that only 1 in 1000 randomly-chosen people would have a DNA profile like the one found in the defendant’s blood and in the blood at the crime scene, so we must conclude that the probability is 99.9% that defendant is guilty.” It is no more the case that $p = .05$ means that we can be ninety-five percent sure that the indicated correlation exists than it is the case that a scarce sample common to the defendant and the crime scene makes for near-certainty that defendant is guilty. Of course it is possible to make a link between the one-in-twenty probability of reaching the observed outcome by chance and the probability of actual cause. Likewise, it is possible to link the “1 in 1000 randomly-chosen people” probability to the probability that defendant was at the crime scene. Doing so, however, involves use of Bayes’ Theorem, which brings new complications and raises the

⁷² See David W. Barnes, *Too Many Probabilities: Statistical Evidence of Tort Causation*, 64 LAW & CONTEMP. PROBS. 191, 208-09 (2001) (there is “no convenient way to translate the .05 p-value into a ninety-five percent confidence that the fact probability is correct,” in part because that value “assumes that the hypothesis is true” and “does not measure whether it is true”) [hereinafter Barnes, *Too Many Probabilities*].

serious problem of assigning a prior probability to the point in issue.⁷³

Now the argument that *Daubert* is leading courts *mistakenly* to require of statistical outcomes the same level of significance that scientists normally require proceeds in this way: This strong scientific bias may be appropriate in the setting of science, but not in the setting of civil litigation. Science observes this strong bias because science can afford it. Science works incrementally and has “forever” to get it right. Here is the way that one modern text in statistics explains the strong scientific bias:

We want α to be small because Type I errors are expensive for the scientific (and the human) enterprise. Suppose, for example, we . . . report in a journal that doses of vitamin B₁₂ increase IQ. That will be an error that we made in good faith because we had no way of knowing that this particular result was a type I error As a result, our readers will alter their behavior, perhaps focusing on a B₁₂ diet while ignoring other avenues (such as reading enhancement programs) that might be effective in raising IQ. Sometime later, perhaps, someone will conduct many experiments and find that vitamin B₁₂ has *no* effect on IQ; that is, they will demonstrate that we had made a Type I error. Our readers who believed our earlier report were done a possibly uncorrectable disservice because they may have ignored other avenues. To prevent any further damage, we would want to find and contact the entire readership of our first report and inform them that our result was mistaken. That is clearly an expensive and difficult (if not impossible) thing to do.

In contrast, the same text continues, we do not need as much protection from Type II errors (β -errors):

Consider an investigator who conducts a single experiment to demonstrate that doses of vitamin B₁₂ increase IQ. After

⁷³ Bayes' Theorem describes the degree to which an item of evidence, when expressed as a datum of known frequency, affects one's prior estimate of the issue on which the evidence bears. For accounts of Bayes' Theorem, see David W. Barnes, *Too Many Probabilities*, at 208-09, and CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE § 7.18 (3d ed. 2003). One problem in utilizing Bayes' Theorem in this setting (indeed *any* setting) is that it is necessary to quantify the *prior* estimate of the probability before using the theorem to find the *new* probability, after taking into account the datum of known frequency. If the *prior* estimate were .05 (very low probability of causation), then applying a statistical finding returned at the conventional level of statistical significance raises the odds to 1:1, meaning equilibrium, or a .5 probability that the degree of cause indicated by the finding (let us say 22 out of 42 instances) is true, or in other words even odds that 22 out of 42 instances of ailment Y in the exposed population exposed were caused by agent X.

the data are collected, she concludes that the null hypothesis should *not* be rejected; that is, she concludes that vitamin B₁₂ has no demonstrable effect on IQ. As a consequence, the investigator will *not* report findings in a journal. Instead, the investigation may be considered an exploration of a blind alley. She had thought vitamin B₁₂ was effective, but apparently it wasn't.

If subsequent research indicates that the conclusion of the original investigation was a Type II error, what is the expense of that error and who bears it? One major expense is the time lost in the original investigation, but the bearer of that expense is the original investigator, not the scientific community at large. There is *no* necessity of informing the community of a previous mistake because there was no report of findings in the first place⁷⁴

Now, the argument continues, our civil justice system differs from science in its goals and social purposes. To start with, our civil justice system does not have unlimited time. We cannot and do not defer decision in the interest of becoming more certain, and correlatively we hold that a lawsuit must reach a conclusion now. Moreover, we must leave the results reached in a lawsuit in place forever. Our legal system cannot be tentative about its conclusions. In a sense, our legal system views "getting it right" as less important than "getting it done." Equally important, our civil justice system is premised on the principle that mistakes in denying relief are as harmful as mistakes in allowing recovery (plaintiff loses when the evidence is equally balanced only because we need a rule to resolve this case).⁷⁵ Thus our civil justice system is neutral as between "false positives" and "false negatives" (α -errors as opposed to β -errors, Type I as opposed to Type II errors). For our civil justice system, *failing* to

⁷⁴ HURLBURT, *supra* note 21, at 196 (acknowledging, however, that Type I and Type II errors can *both* be costly; citing the example of a Type I error in the form of a false report that a drug is effective against AIDS, the author recognizes that this error could have "cruel effects, such as raising false hopes or discontinuing the funding of some other research"; the author also acknowledges, however, that a Type II error in "failing to report a drug that is in fact effective" would "also have cruel results, depriving needy individuals of effective treatment," concluding that there is "no statistical answer" to the question which kind of error is more costly; it is "a matter of complex human judgment").

⁷⁵ Obviously criminal cases are another matter. There our law strongly favors acquittals over convictions if the evidence is in close balance, which is somewhat akin to the idea of preferring false negatives (such as not finding cause when in fact there is cause) over false positives (like finding cause where none exists). Some would argue that the bias of science would be appropriate in discouraging the state's use of thin scientific evidence and grotesquely inappropriate in discouraging the defense use of thin scientific evidence.

find a cause that exists is as bad as finding one that does *not* exist. Hence our civil justice system should not require the level of statistical significance required by science, with its heavy bias against false positives.

These differences between law and science do indeed suggest that we should consider carefully the possibility of accepting results in lawsuits that scientists are not yet prepared to accept. In this paper, I am not prepared to stake out a final position on this issue. Frankly, I'm not sure what the right answer is. However, I do want to address the question how this argument fits with *Daubert*, to talk briefly about who ought to resolve the argument, and to raise some cautionary points.

First, I think the *Daubert* framework can accommodate the view that courts ought to accept scientific evidence that does not satisfy the conventional standard that scientists require. *Daubert* recognizes that the enterprises of law and science differ, and adopts the view that judges must make their own decision, in context and with reference to the needs of the legal system, on admitting or excluding evidence proffered as science. It is true, as some have pointed out in criticizing *Daubert*, that there is language suggesting that courts should be more careful of science than is the professional community that produces science.⁷⁶ But given the more basic premise of *Daubert* that the *law* must judge science for its own purposes, I don't think this language is a serious obstacle to arguments favoring the admissibility of careful studies showing, for instance, a causal connection between ailment Y and agent X even if the results do not satisfy the conventional significance standard.

Second, in this symposium Professor Cohen suggests that a scientist who has evidence indicating, for example, some marginal correlation between agent X and ailment Y, should be able to testify even if she says the results do not satisfy the conventional standard.⁷⁷ In Professor Cohen's formulation, the scientist is seen *explaining* to the jury that science would not accept the indicated conclusion but that the scientist herself might do so for purposes of resolving a question that could not wait. This proposal merits consideration, but it also raises questions. To begin with, and I think Professor Cohen

⁷⁶ See *Daubert*, 509 U.S. at 596-97 (noting that scientific conclusions are "subject to perpetual revision," and science advances through "broad and wide-ranging consideration of a multitude of hypotheses" that can "eventually" be thrown out if wrong, but "[c]onjectures that are probably wrong are of little use" to the law, that must reach a "quick, final, and binding" judgment").

⁷⁷ See Neil Cohen, *The Gatekeeping Role in Civil Litigation and the Abdication of Legal Values in Favor of Scientific Values*, 33 SETON HALL L. REV. 943 (2003).

agrees here, *Daubert* requires *judges* to determine what level of certainty or confidence is high enough to merit consideration by a jury. It is not up to an expert or a jury to decide whether or not to accept evidence that does not satisfy the significance standard. More importantly, telling jurors they can base their verdict on evidence that science would not accept may not be the right thing to do, and it is certainly a strange message: In effect, it says that “you may conclude that defendant caused plaintiff’s cancer even when qualified experts think the case is unproved.” It is even questionable whether a scientist would feel comfortable (or able to comply with the oath required of witnesses) if she testified that while she does not *professionally* accept the indicated conclusion, she might *personally* do for purposes of resolving a lawsuit.

Third, how clear is it that differences between the agendas of science and law justify applying a much more lenient standard to science offered as proof in litigation? It is true that in any one lawsuit there is but one chance to get it right, but it is certainly not true that society as a whole, operating through the legal system, has but one chance. When apparent toxic agent X appears, the system provides many opportunities to resolve the question whether agent X causes ailment Y, and the common-law method of building step by step on experience actually bears some resemblance to the scientific method of moving incrementally and withholding judgment until proof comes that is persuasive. On the other side of the ledger, it is not true that the decision in one case affects only that case. Particularly with medicines and other substances believed to have toxic effect, every court judgment has ripple effects, encouraging or discouraging parallel suits and settlements, and sometimes having legal impacts on later judgments.⁷⁸ Hence errors made in the judicial system, whether favoring claimants or defendants, can produce additional errors as lawyers, claimants, and defendants react to them.

Perhaps equally importantly, it is not really the case that our legal system is “neutral” with respect to errors. In single cases we may be “neutral” as between errors favoring claimants and errors favoring defendants, but we are not neutral *in aggregate* on such points. It is in part because courts recognize the perils of being wrong in huge cases that we have a series of decisions in the federal system that cut back

⁷⁸ For well-known reasons, new claimants are not collaterally estopped by judgments won by the same defendant against prior claimants, and usually new claimants cannot take advantage of collateral estoppel against a defendant stemming from defendant’s prior loss on similar claims. But decisions admitting or excluding scientific proof, or holding it sufficient or insufficient, are likely to have *stare decisis* effects in later suits.

on the use of class suits to resolve mass tort cases.⁷⁹ Even on the supposition that toxic exposure cases are litigated individually, wrongly finding for the plaintiff would have considerable collateral effect, in likely reactions by the defendant and resultant changes in the availability of products. On the supposition that toxic exposure cases are litigated in aggregate fashion, these collateral effects are even clearer and more pronounced.

Finally, I want to suggest that the apparent caution of science may not be quite what it seems. The conventional standard, after all, is just *numbers*—just the product of analysis of the quantification of data. Behind the numbers are more and real uncertainties—the ones that go with designing tests, selecting cohorts, trying to eliminate differences *apart from the factor in issue* that might account for observed differences. Part of the reason science insists on impressive numbers may be the recognition that it is hard or impossible to eliminate confounding variables, and that even promising results might not be replicable. To the extent such apprehensions underlie the insistence on high numbers, the conventions of science are not conservative. And there is an additional factor to consider, which is that scientists (like the rest of us) *want* to be noticed, and have an incentive to maximize the importance of their findings, which suggests that the high conventional standard for statistical significance acts as a counterbalance against self-serving human motivations that are in play not only among lawyers and politicians, but among scientists too, and indeed the whole human species.⁸⁰

⁷⁹ See, e.g., *Ortiz v. Fibreboard Corp.*, 527 U.S. 815 (1999) (disapproving settlement of asbestos claims, largely on the basis of concern over adequate representation); *Amchem Prods. v. Windsor*, 521 U.S. 591 (1997) (refusing to allow certification of nationwide settlement class in asbestos case, largely on basis of concerns over adequate representation and because remedies under consideration cannot be created judicially and require legislative consideration); *In re Rhone-Poulenc Rorer, Inc.*, 51 F.3d 1293 (7th Cir. 1995) (refusing to certify class in suit against maker of blood solids, largely because of reluctance to stake future of defendants on outcome of single trial).

⁸⁰ See, e.g., Lena Williams, *Stalking the Elusive Healthy Diet; In Scientific Studies, Seeking the Truth in a Vast Gray Area*, N.Y. TIMES, Oct. 11, 1995, at C1 (Harvard epidemiologist comments that epidemiology is “a crude and inexact science,” and that “[e]ighty percent of cases are almost all hypotheses,” and that epidemiologists “tend to overstate findings, either because we want attention or more grant money”). See also Robert L. Park, *The Seven Warning Signs of Bogus Science*, 49 CHRON. OF HIGHER EDUC. 21, Jan. 31, 2003, at B20 (as “warning signs,” listing the fact that the scientist “pitches the claim directly to the media,” that she says “a powerful establishment is trying to suppress” her work, that the effect is “at the very limit of detection,” listing “anecdotal” evidence, stressing that the belief has “endured for centuries,” that the discoverer has “worked in isolation,” and that “new laws of nature” are required to understand the discovery).

IV. DAUBERT CAN BE IMPROVED: APPELLATE REVIEW MADE REAL

This paper defends the *Daubert* approach, but one troubling aspect of *Daubert* can and should be fixed. That is the misplaced emphasis on the discretion of trial judges, which appears particularly in the Court's two follow-up decisions in the *Daubert* line. Indeed, it is at least possible to speculate that the sequence in which *Daubert*, *Joiner*, and *Kumho Tire* were decided has much to do with the growth in the emphasis on discretion. The idea is hardly mentioned in *Daubert* itself, but it gained prominence in *Joiner* partly because the Ninth Circuit had adopted an implausible rule that trial judges have *less* discretion to exclude than to admit evidence proffered as science (the antidote was to hold that judges have discretion either way), and it made further gains in *Kumho Tire* because the project in that case entailed explaining how the standard for science could be applied usefully to experiential expertise (where a measure of discretion seems essential if the scheme is going to work at all).⁸¹

I concur with commentators who say *Daubert* should be implemented by inviting appellate courts to take a close look at rulings by a trial court admitting or excluding evidence offered as science.⁸² Having become accustomed to the refrain among federal appellate courts that they accord deference to trial court decisions applying *Daubert*, I was surprised to learn that nine states and the District of Columbia instruct appellate courts to review rulings admitting or excluding evidence presented as science by applying a *de novo* standard.⁸³ Some decisions apply this rigorous standard only to

⁸¹ In describing the gatekeeping function in *Daubert*, it is singular that the Court made *no reference* to discretion. By my count, the Court mentioned discretion ten times in describing this function in *Joiner* (making many additional passing references to the term), and six times in the same context in *Kumho Tire*. See *Daubert*, 509 U.S. at 594 (mentioning that its standards are “flexible”); *General Elec. Co. v. Joiner*, 522 U.S. 136, 141-47 (1997) (repeatedly stressing discretion); *Kumho Tire*, 526 U.S. at 151 (repeatedly stressing discretion, and pointedly saying that judge “must have the same kind of latitude in deciding *how* to test an expert’s reliability” that it has “when it decides *whether or not* that expert’s relevant testimony is reliable”) (emphasis in original).

⁸² See David L. Faigman, *Appellate Review of Scientific Evidence Under Daubert and Joiner*, 48 HASTINGS L.J. 969 (1997) (for ordinary decisions to admit evidence, where preliminary facts “depend on the testimony of witnesses,” appellate deference is warranted, but scientific evidence is “quite different,” and the trial judge is not in a “preferred position” in evaluating it). See also Richard D. Friedman, *Squeezing Daubert Out of the Picture*, 33 SETON HALL L. REV. 1047, 1065 (2003) (appellate courts should play more of a role in reviewing *Daubert* issues than *Joiner* suggests).

⁸³ I looked at modern decisions from all fifty states, most of which endorse an abuse-of-discretion standard. But a *de novo* standard has been adopted in Arizona, Florida, Kansas, Maryland, Minnesota, New Jersey, Oregon, Oklahoma, and Washington, as well as the District of Columbia. See generally *Goeb v. Tharaldson*, 615

the basic question whether the theory and method are valid—or in *Frye* terms, to the proxy question whether they are generally accepted—and leave related questions like “helpfulness” under FRE 702 for the trial judge to decide under an abuse-of-discretion standard.

There are four reasons for preferring a more exacting standard in review.

First, issues relating to the validity of theories and techniques transcend the facts of individual cases. This observation applies, for example, to the question whether DNA profiling can reliably identify a blood or fluid sample as having very likely come from one person or another (the validity of the theory), to the question whether particular methods of analysis (such as RFLP, PCR, and STR) accurately measure the attributes of blood or fluid, and whether a particular laboratory protocol adequately guards against missteps and laboratory error.⁸⁴ It applies to the question whether proffered

N.W.2d 800, 814-15 (Minn. 2000) (whether proffered expertise satisfies state’s *Frye-Mack* “general acceptance” standard “is a question of law that we review *de novo*,” but questions of “foundational reliability” are reviewed under “abuse of discretion” standard, as are matters of witness qualification); *Kuhn v. Sandos Pharm. Corp.*, 14 P.3d 1170, 1179 (Kan. 2000) (adopting *de novo* standard of review for proof of medical causation); *Jennings v. Baxter Healthcare Corp.*, 14 P.3d 596 (Or. 2000) (rejecting argument that appellate court should accord deference to trial court’s decision on scientific validity, and concluding that the issue is reviewed as for “errors of law”); *Hadden v. State*, 690 S.2d 573, 578 (Fla. 1997) (review of *Frye* issues is *de novo*); *State v. Harvey*, 699 A.2d 596, 619 (N.J. 1995) (in applying *Frye* standard, question whether scientific community generally accepts a method or test “can transcend a particular dispute,” and to the extent that *Frye* focuses on “issues other than a witness’s credibility or qualifications, deference to the trial court is less appropriate”); *Taylor v. State*, 889 P.2d 319, 331 (Okla. 1995) (decision by trial court to admit novel scientific evidence should be subject to “an independent, thorough review,” and appellate court should “not simply ask whether an abuse of discretion was committed”); *State v. Tankersley*, 956 P.2d 486, 464 (Az. 1994) (rejecting *Daubert* and staying with *Frye*, and announcing that *Frye* issues are subject to *de novo* review); *Schultz v. States*, 664 A.2d 60, 64 (Md. App. 1994) (question of reliability of scientific technique “does not vary according to the circumstances of each case,” so it is inappropriate to apply abuse of discretion standard on review); *State v. Cauthron*, 846 P.2d 502, 505 (Wash. 1993) (stating that court would “review the trial court’s decision to admit or exclude novel scientific evidence *de novo*”); *United States v. Porter*, 618 A.3d 629, 634 (D.C. 1992) (questions of general acceptance of new scientific techniques invite court “to establish the law of the jurisdiction for future cases,” so court would “engage in a broad review”).

⁸⁴ The initials cited above refer to three of the more common methods for conducting DNA profiling. RFLP refers to “restriction fragment length polymorphism,” and it is the first broadly useful approach that made its way into courtroom use. PCR refers to “polymerase chain reaction,” a later development that allowed small samples to be “extended” so that the inevitable consumption of such materials in laboratory testing did not destroy the whole sample. The drawback of PCR was that it could only extend some of the attributes of the original sample, so

statistical proof should satisfy the standard that scientists would require, to the question whether differential diagnosis can or cannot, standing alone, prove specific cause, to the question whether animal studies of any particular drug or chemical can prove causation in the human population, and to the question whether similarities between the chemical structures of a particular drug or other substance, on the one hand, and some other agent known to cause certain consequences, such as disease, on the other hand, can prove causation. Questions of this magnitude need steadier guidance than the abuse-of-discretion standard provides, and the answers that courts reach should be applied in similar cases, rather than left to vary with the differing views of trial judges exercising discretion.

Second, appellate courts are better situated than trial courts to resolve such questions. To start with, three or more minds are likely to do better than one in appraising such technical issues. And appellate review goes forward in a setting less subject to severe schedule pressures. Furthermore, *Daubert* issues are likely to benefit from thorough appellate briefings. And appellate review can involve consultation with technical materials and expert advice by means of amicus briefs or affidavits, or even live testimony. Reviewing courts can even take judicial notice of technical books, articles, and other materials.⁸⁵ Some sense of proportion is clearly warranted: It is one thing to supplement the arguments and briefs of counsel with references to additional material, and another thing to decide the case on grounds never considered by the lawyers who briefed the case without giving them any opportunity for input.

Third, the *Daubert* standard needs elaboration in the variety of settings in which it is to apply, and trial judges need help and guidance beyond that provided by the standards themselves. The problem of causation in toxic tort cases is a prime area in which appellate courts could play useful roles, and in which trial courts clearly want and need guidance.

There is of course one countervailing concern, and that is that

testing could not be as extensive. STR, or “short tandem repeats,” is a still more recent development. *See generally Shreck*, 22 P.3d at 73 (describing these techniques).

⁸⁵ FRE 201 governs only judicial notice of “adjudicative” facts, and most technical material that might be noticed in this setting involves “evaluative” facts utilized by courts in their attempts to formulate wise rules of law. The fact that FRE 201 does not cover evaluative facts does not mean they cannot be noticed. Instead, the omission from coverage simply means that judicial notice of evaluative facts is not regulated by any formal rule. *See State v. Jones*, 922 P.2d 806, 809 (Wash. 1996) (*de novo* review includes “sources outside the record such as scientific literature, law articles, and the decisions of other jurisdictions”) (applying state’s *Frye* standard).

the *Daubert* standard is both wide-ranging and case-specific. It is wide-ranging now, if it was not when *Daubert* was decided, because amended FRE 702 indicates that judges are to consider “principles and methods” and the sufficiency of underlying “facts or data,” and also the question whether the expert “has applied the principles and methods reliably to the facts,” and *Kumho Tire* makes it clear that the focus is the “task at hand,” as Professors Denbeaux and Risinger remind us.⁸⁶ To the extent that the admissibility decision actually focuses, for example, on the question whether a particular laboratory protocol was or was not followed in the case at hand, or on the question whether a particular lapse or discrepancy in the data materially affected the outcome, some degree of deference to the decision of the trial judge is in order. It is with larger questions, including those of theory and technique, and the appropriateness of the technique to the issue being decided, that closer scrutiny is warranted.

CONCLUSION

The Court’s decision in *Daubert* changed the relationship between law and science. Critics have argued that judges cannot act constructively in the way that *Daubert* envisions, but there are good reasons to think that indeed judges can rise to the task. Critics have argued as well that *Daubert* has led to the exclusion of too much scientific evidence, particularly in toxic tort cases, and have adopted restrictive rules that are out of place and, in the federal system, infringe on the *Erie* doctrine. But *Daubert* does not require adoption of such rules, and the cases show that courts are in fact working hard in very challenging areas to achieve appropriate outcomes in appraising science. The *Erie* doctrine is not offended by federal efforts to implement a sufficiency standard.

Critics have also argued that *Daubert* misconceives science, and the relationship between law and science. But the dualism visible in *Daubert*’s account of science is also visible in the accounts of philosophers and historians of science, and the task is to reconcile notions of objectivity and subjectivity in scientific undertakings. The truly difficult question whether scientific evidence proffered in civil

⁸⁶ See amended FRE 702 (described in the text accompanying note 51, *supra*); see also *Kumho Tire*, 526 U.S. at 154 (stressing that the question for the court is *not* “the reasonableness in general” of a particular technique, but “the reasonableness of using such an approach . . . to draw a conclusion regarding the particular matter to which the expert testimony was directly relevant”) (emphasis in original); Mark P. Denbeaux & Michael Risinger, *Kumho Tire and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 SETON HALL L. REV. __ (upcoming in Fall 2003).

cases should achieve a level of certainty that scientists themselves would require has not yet been resolved, but *Daubert* leaves room either to require that level or to admit scientific evidence on a lesser showing of significance. Which choice should be made here remains open to debate.

In the federal system, reviewing courts speak highly of the discretion that trial judges have in applying the *Daubert* standards, but a handful of states follow a different approach in allowing reviewing courts to appraise claims of error in applying *Daubert* on a *de novo* basis. These courts are doing the right thing, as trial judges need more extensive appellate guidance in handling science in civil cases under the *Daubert* standard.