MIND THE ANALYTICAL GAP!

TRACING A FAULT LINE IN *DAUBERT*[1]

SUSAN HAACK[2]

> "It may be said that the difference is only one of degree. Most differences are, when nicely analyzed." — Oliver Wendell Holmes[3]

As I learned on a recent visit to Porto Alegre, when a Brazilian has a problem he'll say, "tenho um abacaxi para resolver" — "I have a pineapple to resolve." It's a charming idiom; but, sadly, the problem on my mind today isn't just prickly, but also hugely messy—it's a big, tangled knot of questions, not nearly as compact as a pineapple. How did it come about—given that *Daubert* describes the old *Frye* Rule as "austere," and the new régime introduced with Federal Rule of Evidence [FRE] 702 as more flexible, more favorable to the admission of expert testimony[4]—that, at least in civil cases, the effect of *Daubert* seems to have been to make the admission of expert testimony *more* difficult, rather than *less*, and even to shift admissibility closer to sufficiency? Is there something about the character of relevance and reliability, the key concepts of *Daubert*, that explains this development? How did these two concepts get built into *Daubert* to begin with? Are these post-*Daubert*

---

1. I. Goodman Cohen lecture, Wayne Law School, October 2015. © 2015 Susan Haack.

2. Distinguished Professor in the Humanities, Cooper Senior Scholar in Arts and Sciences, Professor of Philosophy, and Professor of Law at the University of Miami. B.A., M.A., B.Phil., Oxford; Ph.D., Cambridge. Professor Haack's most recent books are EVIDENCE MATTERS: SCIENCE, PROOF AND TRUTH IN THE LAW (2014), PERSPECTIVAS PRAGMATISTAS DA FILOSOFIA DO DEREITO (2015), and LEGALIZZARE L'EPISTEMOLOGIA (2015).

3. Rideout v. Knox, 19 N.E. 390, 392 (Mass. 1889) (Holmes, J. majority opinion). The distinction at issue in this case was the permissible height of boundary fences; but Holmes's point is of course quite general.

4. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 588-89 (1993) ("Daubert I").

developments to be welcomed? Or are they not only unintended but also undesirable consequences of a potentially dangerous fault line?

How do you resolve a pineapple? I suppose, the same way you eat an elephant—one bite at a time! And how do you disentangle a big, messy knot of questions? Well, the only way I know is patiently to tug out one strand at a time. So I'll start by sketching enough of the pre-history of *Daubert* to suggest how the concepts of relevance and reliability came to play their central role (§1). Then I'll explore some complexities of these concepts, most importantly their gradational character (§2). Next, I'll argue that the mismatch between the gradational concepts of relevance and reliability and the categorical concept of admissibility presents courts with a problem about the degree of relevance and degree of reliability to require; and show by means of a sampling of rulings on *Daubert* issues that, while some courts have maintained a clear distinction, others have set the bar of relevance and/or of reliability so high as to blur the line between admissibility and sufficiency (§3). And finally, tugging at the normative strands in my knot of problems, I'll try to explain why I find this elision of admissibility into sufficiency disturbing (§4).

1. HISTORICAL PRELIMINARIES

Of course, the concept of relevance is hardly new to legal thinking. It figures prominently, for example, in two famous volumes on evidence law published long before *Daubert*: James Thayer's celebrated *Preliminary Treatise on Evidence at the Common Law*, published in 1898,[5] and Dean McCormick's influential textbook, *Handbook of the Law of Evidence*, published in 1954.[6]

An entry for "relevancy" in the index of Thayer's treatise takes you to a substantial chunk of text. Relevancy, Thayer tells us, is a "fundamental conception," since a presupposition of any rational system of legal proof is that evidence which is not relevant should not be admitted.[7] That relevant evidence *should* be admitted, he goes on, is also a fundamental principle; but this time subject to many qualifications and exceptions because, over the course of its history, our legal system has

5. JAMES BRADLEY THAYER, A PRELIMINARY TREATISE ON EVIDENCE AT THE COMMON LAW (1898).

6. CHARLES T. MCCORMICK, HANDBOOK OF THE LAW OF EVIDENCE (1954).

7. THAYER, *supra* note 5, at 264–65.

constructed an elaborate set of rules excluding certain kinds of relevant evidence, along with various exceptions to those rules.[8]

An entry for "relevancy" in the index of McCormick's textbook takes you to an entire chapter on the subject. Citing Thayer at some length, McCormick endorses his two "fundamental principles," which he summarizes like this: "the first ground of exclusion should be want of probative value"; and, "if evidence is logically probative, it should be received unless there is some distinct ground for refusing to hear it."[9] He then proposes an understanding of relevancy as "the tendency of the evidence to establish a material proposition" for which it is offered;[10] and cashes out this "tendency" in terms of the evidence's raising the probability of the truth of the proposition concerned and this, in turn, as its raising the chances that the proposition is true.[11] He also explains what he takes to be the chief grounds that would justify the exclusion of relevant evidence—as the subtitle of his chapter says: "Time, Prejudice, Confusion, and Surprise."[12]

There is *no* index entry for "reliability," however, in either book. True, Thayer's treatise includes a substantial chapter on the so-called "Best Evidence" principle,[13] which presumably has *something* to do with reliability. But, after tracing its long history and its many supposed applications, Thayer argues that the Best Evidence idea should be understood, not as a precise legal rule, but as "a large moral principle":[14] "that always, morally speaking, the fact that any given way of proof is all that a man has, must be a strong reason for receiving it. . . ," and that a party's not producing the best evidence it could "afford[s] strong ground of suspicion."[15] McCormick's book also includes a brief section on the Best Evidence principle,[16] citing Thayer;[17] but this serves merely as the introduction to a long discussion of the ramifications of the "original

8. *Id.* at 265–66. (It strikes me that there is something more than a little odd about acknowledging that the admissibility of relevant evidence is a "fundamental principle," and then suggesting that this principle needs hedging to accommodate the mesh of exclusionary rules our legal system has happened to arrive at. But I can't pursue that thought here.)

9. MCCORMICK, *supra* note 6, at 314.

10. *Id.*at 315–16.

11. *Id.* at 317–18. This explanation in terms of chances makes it clear that McCormick is taking degrees of proof to be mathematical probabilities.

12. *Id.* at 314 (subtitle), 315–21 (text).

13. THAYER, *supra* note 5, at 484–507.

14. *Id.* at 505.

15. *Id.* at 507.

16. MCCORMICK, *supra* note 6, at 408–09.

17. *Id.* at 408.

document" rule.[18] McCormick doesn't mention, however, that Thayer had resisted the idea that this, or any, specific rule is an application of that more general principle.

Most immediately to the present purpose is McCormick's chapter on "Experimental and Scientific Evidence,"[19] in which he contrasts *Frye v. United States* (1923)[20] with *McKay v. State* (1950).[21] He describes *Frye* simply as the first case in which a court faced the question of the admissibility of lie-detector evidence; and *McKay* as the first case in which a court faced the question of the admissibility of the results of the Harger breath test.[22] And he emphatically endorses the approach taken in *McKay*—where the court had ruled that the objection that this expert testimony was not generally accepted went to its *weight*, not its *admissibility*[23]—over the approach taken in *Frye*. The "general scientific acceptance" test proposed in *Frye* as a mark of sufficient "accuracy" (apparently, reliability) to warrant the admission of novel scientific testimony, McCormick avers, is misconceived; this is "*a proper condition upon the court's taking judicial notice of scientific facts, but not a criterion for the admissibility of scientific evidence.*"[24] In short, McCormick's position is that the admissibility of expert testimony should require *only* relevancy, *not* reliability.

In 1954, when McCormick's text was first published, *Frye* had been cited in only a relatively few rulings—and usually, as McCormick's brief description suggests, as a precedent for excluding lie-detector evidence.[25] By 1975, however, when the Federal Rules of Evidence were ratified, the *Frye* "general acceptance" test had been adopted in many jurisdictions across the country,[26] by this time as a rule governing the admissibility of

18. *Id.* at 409–25.

19. MCCORMICK, *supra* note 6, at 359-83.

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

21. McKay v. State, 155 Tex. Crim. App. 416, 235 S.W.2d 173 (1950).

22. The phrase refers to an early precursor of the modern breathalyzer, a roadside breath-testing machine, the "drunkometer," invented by Rolla Neil Harger in 1931. Douglas Martin, Obituary, *Rolla N. Harger Dies. Invented Drunkometer*, N.Y. TIMES, Aug. 10, 1983, at B6.

23. *McKay*, 235 S.W.2d at 175.

24. MCCORMICK, *supra* note 6, at 363 (emphasis added).

25. *Id.* (discussing lie-detector tests); *see* David L. Faigman, Elise Porter & Michael J. Saks, *Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence*, 15 CARDOZO L. REV. 1799, 1808 n.25 (Apr. 1994) (explaining that "*Frye* was not cited by a single other court, federal or state, for a decade. During the first quarter century after its publication, *Frye* was cited in eight federal cases and five state cases.").

26. Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 COLUM. L. REV. 1197, 1228–29 (1980); Faigman, Porter & Saks, *supra* note 25, at 1808 n.25 (clarifying that "[d]uring its second quarter century,

novel scientific testimony generally. But FRE 702—which set the federal standard of admissibility for all expert testimony, including all scientific testimony, whether new or not—made no reference either to *Frye* or to general acceptance. It read, simply:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.[27]

Had FRE 702 superseded *Frye*? Courts, and legal scholars, were divided: some took the new Rule to reject *Frye* and endorse a pure relevancy approach like McCormick's; others thought that, like *Frye*, Rule 702 implicitly required some indication of reliability.[28] In 1985, when Judge Becker wrote his ruling in *U.S. v. Downing*,[29] this controversy remained unresolved.

Mr. Downing had been convicted, solely on the basis of eyewitness testimony that he was the perpetrator, of wire fraud and interstate transportation of stolen property.[30] The legal issue on appeal was whether the district court had erred in excluding the expert in the psychology of perception and memory whom the defense wanted to introduce to testify as to potential flaws of the eyewitness testimony in the case.[31] Ruling that this *had* been a legal error, Judge Becker vacated Downing's conviction and remanded the case for an *in limine* hearing on the admissibility of the proffered psychological evidence. Given that no fewer than twelve eyewitnesses, some of whom had talked with him for as long as forty-five minutes—and not in stressful circumstances, but in the course of routine business dealings—had testified that it was indeed Downing who had passed himself off as the Rev. Claymore and as

---

[*Frye*] was cited fifty-four times in federal cases and twenty-nine times in state cases. By the 1980s, it was being cited as much each year as it had been in its first fifty years added together. What seems apparent from counting case citations is that judicial interest in the *Frye* test did not pick up until a few years before the promulgation of the Federal Rules of Evidence and thereafter, no doubt stimulated by the drafting and adoption of the Federal Rules themselves.").

27. FED. R. EVID. 702 (1975) (amended 2011).

28. The controversies are described in United States v. Downing, 753 F.2d 1224, 1232–33 (1985); Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 587 n.5 (1993) ("Daubert I").

29. *Downing*, 753 F.2d. at 1224.

30. *Id.* at 1226–27.

31. *Id.* at 1226.

representing the Universal League of Clergy,[32] you might be excused for feeling, as Judge Dumbauld suggests in his concurrence, that Judge Becker's substantial, closely-argued ruling is something of a sledgehammer to crack a nut;[33] nonetheless, it is, in Judge Dumbauld's picturesque phrase, "a minor magnum opus of jurisprudential virtuosity."[34]

Noting approvingly that some courts had recently admitted such expert psychological testimony under what has become known as the "helpfulness" provision of FRE 702,[35] observing there are numerous problems with the *Frye* test,[36] that it was doubtful whether *Frye* survived under the more liberal standard of FRE 702,[37] and that other courts had "focused on reliability as a critical element of admissibility,"[38] Judge Becker proposes that, when determining the admissibility of novel scientific evidence, judges should consider:

> (1) the soundness and reliability of the process or technique used in generating the evidence;
>
> (2) the possibility that admitting the evidence would overwhelm, confuse, or mislead the jury; and

---

32. *Id.* at 1226–27.

33. In fact, given that he comments that the circumstances in which the eyewitnesses met with "Rev. Claymore" weren't of the kind in which the expert would testify that mistakes are common—there was neither stress nor cross-racial identification involved—Judge Becker himself seems to anticipate that the proffered expert evidence might in due course be ruled inadmissible on these grounds. *Downing*, 753 F.2d at 1242.

34. *Downing*, 753 F.2d at 1244 (Dumbauld, J. concurring). One has to wonder whether Judge Dumbauld was aware how marvelously oxymoronic this backhanded compliment is, or how much it sounds like a line from one of W. S. Gilbert's libretti!

35. *Id.* at 1230–31. Recall that, as it read in 1975, FED. R. EVID. 702 spoke of expert testimony that will "assist" the finder of fact. Now (2015) it says "help."

36. *Downing*, 753 F.2d at 1235–37. *Frye* is both vague and conservative, Judge Becker argues; moreover, it may exclude reliable evidence that is not yet generally accepted.

37. "[T]he Rule 702 standard usually favors admissibility. . . ." *Id.* at 1229.

38. *Id.* at 1238 (citing State v. Temple, 273 S.E.2d 273, 280 (1981); State v. Kersting, 623 P.2d 1095, 1101 (1981); D'Arc v. D'Arc, 385 A.2d 278 (1978) (providing examples that suggest that scientific methods not yet generally accepted may be admitted if they are shown to be reliable); and United States v. Franks, 511 F.2d 25, 33 n.12 (6th Cir. 1975) as an example of a court suggesting that "general acceptance" is "nearly synonymous with reliability).

> (3) the proffered connection between the scientific research or test result to be presented, and particular disputed factual issues in the case.[39]

Spelling out the first clause, Judge Becker includes a substantial paragraph suggesting how a court might go about determining the "soundness and reliability" of novel scientific testimony, listing factors that might appropriately be considered in the "flexible" inquiry he believes is needed.[40] Degree of acceptance in the relevant community might still be considered, he writes;[41] but where scientific testimony has no such established track record, courts might look to the relation of the theory or technique to more established kinds of scientific analysis;[42] to the existence of a specialized literature;[43] to whether the theory or technique has been subjected to scientific scrutiny;[44] and, if known, to its error rate.[45]

The second clause of Judge Becker's three-part test simply echoes FRE 403, excluding relevant evidence if its probative value is outweighed by the danger that it will waste time, or confuse or mislead the jury.[46] His other two clauses, however, are less familiar—or at least, in 1985 they *were* less familiar. The third amplifies the requirement that admissible evidence be relevant to factual issues in the case, as it applies to scientific testimony specifically; and the first represents a decisive step away from the pure relevancy approach favored by McCormick and others. Nearly a decade later, both of these ideas will be found in *Daubert*.

Daubert v. Merrell Dow Pharmaceuticals[47] was in most respects a routine Bendectin case, indistinguishable from the many other cases alleging that this morning-sickness drug caused birth defects in the children born to women who took it.[48] So why did the Supreme Court grant certiorari?—Because, in ruling the plaintiffs' expert testimony inadmissible, the district court in *Daubert* had relied on the "general

---

39. *Downing*, 753 F.2d at 1237.

40. *Id.* (citing 3 J. WEINSTEIN & M. BERGER, WEINSTEIN'S EVIDENCE §702[03], 702–19 (1981)).

41. *Downing*, 753 F.2d at 1238.

42. *Id.*

43. *Id.*

44. *Id.* at 1239.

45. *Id.*

46. FED. R. EVID. 403 (1975) (amended 2011).

47. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993) ("Daubert I").

48. *See generally* MICHAEL D. GREEN, BENDECTIN AND BIRTH DEFECTS: THE CHALLENGES OF MASS TORTS SUBSTANCES LITIGATION (1996); JOSEPH SANDERS, BENDECTIN ON TRIAL: A STUDY OF MASS TORT LITIGATION passim (1998).

acceptance" test and, in affirming this exclusion, the court of appeals had specifically cited *Frye*[49]—which, however, had up till then been used in criminal trials rather than in civil cases.[50] This provided the Supreme Court with an opportunity to settle once and for all whether *Frye* had, or hadn't, been superseded with the passage of FRE 702. "We granted certiorari [in *Daubert*] in light of sharp divisions among the courts regarding the proper standard for the admission of expert testimony."[51]

Reading Justice Blackmun's ruling in *Daubert*, I for one—and, judging by a comment in his ruling in *Paoli*,[52] Judge Becker for another—have a distinct sense of *déjà vu*. The *Frye* Rule *had* been superseded, the *Daubert* Court ruled;[53] and FRE 702 was more hospitable to the admission of expert scientific testimony than *Frye*.[54] But this didn't mean that the FRE abrogated courts' gatekeeping responsibility.[55] On the contrary, Justice Blackmun argued (as Judge Becker had done), FRE 702 required that courts screen proffered expert testimony both for its "fit" to the factual issues in the case, and for its reliability. "[U]nder the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."[56] (In this context, citing the Advisory Committee Notes on FRE 602, Justice Blackmun refers to the common law's insistence on "'the most reliable sources of information.'"[57] Ironically, the Advisory Committee was quoting McCormick on "best evidence" [58]—but hadn't mentioned that he had emphatically *rejected* the idea that reliability should be required for expert testimony to be admissible.)

Again like Judge Becker, Justice Blackmun finds both of these requirements, relevance and reliability, implicit in the "helpfulness" clause of FRE 702. As he amplifies the relevancy requirement, he

---

49. Daubert v. Merrell Dow Pharm., Inc., 727 F. Supp. 570, 572 (1989); Daubert v. Merrell Dow Pharm., Inc., 951 F.2d 1128, 1129–30 (9th Cir. 1991).

50. But, I should add, not quite exclusively so. *See* Barrel of Fun v. State Farm Fire & Cas. Co., 739 F.2d 1028 (5th Cir.1984); Christopherson v. Allied Signal Corp., 902 F.2d 362 (5th Cir.1990), *superseded by* 939 F.2d 1106 (5th Cir. 1990), *and cert. denied*, 503 U.S. 912 (1992). *See also* Kenneth J. Cheseboro, *Galileo's Retort: Peter Huber's Junk Scholarship*, 42 AM. U. L. REV. 1637 (1993).

51. *Daubert I*, 509 U.S. at 585 (citations omitted).

52. *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 742 (3rd Cir. 1994) (noting the similarity between *Daubert*'s list of indicia of reliability, and Judge Becker's own in *Downing*).

53. *Daubert I*, 509 U.S. at 589.

54. *Id.* at 588–89, and n.6.

55. *Id.* at 589.

56. *Id.*

57. *Id.* at 592 (citing FED. R. EVID. 602 advisory committee's note on FED. R. EVID. 602, 28 U.S.C. App., p.755).

58. MCCORMICK, *supra* note 6, at 19.

borrows Judge Becker's word, "fit," and quotes his explanation;[59] and, after articulating the reliability requirement, again like Judge Becker, he goes on to suggest how judges might go about assessing the reliability of proffered expert testimony. Here, however, Justice Blackmun begins with an idea not found in *Downing*: the subject of an expert's testimony must be "scientific . . . knowledge," he writes; and this "establishes a standard of evidentiary reliability."[60] "Knowledge," he continues, "connotes more than subjective belief or unsupported speculation"; and "scientific" "implies a grounding in the methods and procedures of science."[61] Hence his insistence that the focus be on a proffered expert's methodology, rather than his conclusions;[62] and hence, also, his first indicium of reliability (can it be and has it been tested?), which relies on a quasi-Popperian misconception of the supposed "scientific method."[63] But the other factors on Justice Blackmun's flexible list—peer review and publication, known or potential error rate, widespread acceptance in the relevant community[64]—are very reminiscent of Judge Becker's recommendation that, unless a theory has a track record of acceptance, judges should ask whether it has been subjected to scientific scrutiny, whether there is an established scientific literature on the subject, and what the error rate of a scientific technique might be.[65]

So the Supreme Court ruled that *Frye* was superseded;[66] and then imposed a requirement that expert testimony be not only relevant and "fit" the facts at issue, but also reliable. Given that reliability first entered the picture as a criterion for admissibility, rather than as a factor going to weight, in *Frye*, this is more than a little ironic. And this is only one of many ironies in *Daubert*: among which I would also include the vain hope that Karl Popper's falsificationist philosophy of science—which amounts, in the end, to nothing but a thinly-disguised form of

---

59. In the context of his discussion of the "fit" requirement, Justice Blackmun refers specifically to *Downing*. *Daubert I*, 509 U.S. at 591–92.

60. *Daubert I*, 509 U.S. at 589–90 n.8. The adjective "evidentiary" signals that this is a new, legal concept, not quite to be identified with the ordinary sense of "reliable." The ellipses signal that Justice Blackmun has quietly dropped "technical, or other specialized" from the text of FRE 702. FED. R. EVID. 702.

61. *Daubert I*, 509 U.S. at 590.

62. *Id.* at 595.

63. *Id.* at 593; *see also* Susan Haack, *Trial and Error*, 95 AM. J. PUB. HEALTH S66 (2005), *reprinted in* EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 104 (2014); Susan Haack, *Federal Philosophy of Science: A Deconstruction—And a Reconstruction*, 5 N.Y.U. J. L. & Liberty 394 (2010), *reprinted in* EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 122 (2014).

64. *Daubert I*, 509 U.S. at 593–94.

65. *See supra* notes 41–45 and accompanying text.

66. *Daubert I*, 509 U.S. at 589.

skepticism[67]—might serve as a criterion of the reliability of scientific testimony;[68] and the suggestion that peer-reviewed publication is another indicator of reliability[69]—when much of the substantial body of peer-reviewed literature indicating that Bendectin was not teratogenic seems, if Judge Bernstein's ruling in *Blum v. Merrell Dow* (1996) is to be believed,[70] to have been based on work conducted by scientists employed, or supported, by Merrell Dow, and some of which seems to have been seriously flawed.[71] Most to the present purpose, though, is the startling mismatch between the plain statement that *Frye* is an "austere standard," now superseded by the new, more liberal FRE 702,[72] and the imposition of what has often turned out to be significantly *more* rigorous judicial screening of expert testimony than before.

Merrell Dow had suggested that abandoning *Frye* would result in "a 'free-for-all' in which befuddled juries are confounded by absurd and irrational pseudoscientific assertions";[73] but Justice Blackmun reassured them: the adversary system provides sufficient protection against this danger. "[C]ross-examination, presentation of contrary evidence, and . . . instruction on the burden of proof," he wrote, are "the traditional and appropriate means of attacking shaky but admissible evidence."[74] Presumably, the reason he took the new standard he articulated in *Daubert* to be more liberal than the old *Frye* Rule was that general acceptance was no longer required—scientific ideas that were not yet generally accepted could nonetheless be admitted if other *Daubert* factors were satisfied.[75] It seems safe to say, however, that in practice, at

---

67. *See* SUSAN HAACK, *Just Say "No" to Logical Negativism* (2011), in PUTTING PHILOSOPHY TO WORK: INQUIRY AND ITS PLACE IN CULTURE 179 (2008; 2nd ed. 2013); *Federal Philosophy of Science*, *supra* note 63.

68. *Daubert I*, 509 U.S. at 593.

69. *Id.*

70. Blum v. Merrell Dow Pharm., Inc., 33 Phila. Cnty. Rep. 193, 230 (1996).

71. For more details, *see infra* note 88 and accompanying text; Susan Haack, *Peer Review and Publication: Lessons for Lawyers*, 36 STETSON L. REV. 789 (2007), *reprinted in* EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 156 (2014).

72. *Daubert I*, 509 U.S. at 589.

73. *Id.* at 595–96.

74. *Id.* at 596. Of course, as we have seen (and as Merrell Dow must surely have known) it wasn't true that, until *Daubert*, *Frye* had kept "junk science" out of civil cases; it had almost always been confined to criminal cases.

75. Of course, *Daubert* applied only to federal courts; and when the Florida Supreme Court decided that Florida should not adopt *Daubert*, but stick with *Frye*, it took pride in maintaining its (supposedly) more rigorous régime. "Despite the federal adoption of a more lenient standard . . . we [Florida] have maintained the higher standard of reliability as dictated by *Frye*." Brim v. State, 695 So. 2d. 268, 271-72 (1997). The present situation in Florida is far from clear. Under the Florida constitution, procedural changes are the province, not of the legislature, but of the Florida Supreme Court. FLA. CONST. art. 5 §2.

least in civil cases, *Daubert* has proven much more demanding than, apparently,[76] Justice Blackmun anticipated.

In part, this is because—given that it interprets FRE 702, which applies to *all* expert testimony, scientific or not, novel or not—*Daubert* is much broader in scope than *Frye*, which applies only to novel scientific testimony. In part, also, it's because, if a party is obliged to show that its expert testimony passes muster under *all* the *Daubert* factors, it could face a significantly harder task under *Daubert* than it would have under *Frye*. But in significant part, too, as I shall argue, it is due to certain key complexities of those central *Daubert* concepts, relevance and reliability—to the potential for conceptual slippage that I described as a "fault line" in *Daubert*.

2. CONCEPTUAL COMPLEXITIES

Relevance is a two-place relation—a relation between some proposition, claim, evidence, testimony, consideration, point, fact, etc.,

---

However, in 2013 the Florida legislature voted to adopt *Daubert*. 2013 FLA. LAWS 107, FLA. STAT. ANN. §§ 90.702 & 90.704 (West 2014). Then, in *Hernandez v. State*, the Supreme Court of Florida acknowledged that this move to *Daubert* was intended "to tighten the rules for admissibility of any expert opinion"; but did not unambiguously endorse the change. Hernandez v. State, 2015 WL 5445655, at *24 (Sept. 17, 2015). Granted, the ruling continues, in *Daubert* the U.S. Supreme Court had criticized *Frye* as too restrictive; nonetheless, even under *Daubert* general acceptance in the relevant community remains one factor among several. *Id.* at *23–24. Shortly afterwards, the Daily Business Review reported that the Florida Justice Association, a statewide advocacy group for plaintiffs' attorneys, is urging that Florida "go back" to *Frye*. Julie Kay, *Panel Urges Going Back to* Frye *Standard for Expert Testimony*, DAILY BUS. REV., Oct. 20, 2015, at A1. As of mid-November 2015, the Florida Bar seemed to be divided on the issue. *Compare* Wayne Hogan, *Daubert—A Burdensome Solution in Search of a Problem*, FLA. B. NEWS, Nov. 15, 2015, at 10 *with* David A. Jones, *Frye v. Daubert—Time to Bring Florida Into the 21st Century*, FLA. B. NEWS, Nov. 15, 2015, at 10. But in December 2015, the Board of Governors of the Florida Bar Association voted to approve the bar committee's recommendation that Florida continue to use *Frye*. FLORIDA BAR BOARD OF GOVERNORS, REGULAR MINUTES 8 (Dec. 4, 2015), http://www.floridabar.org/TFB/TFBResources.nsf/Attachments/0E799A221F4FB5B385257F4D0077044A/$FILE/Regular%20Minutes,%20December%204,%202015%20meeting.pdf.

76. Why "apparently"?—because it might be thought that the *Daubert* Court's rhetoric about FED. R. EVID. 702's being more flexible and hospitable to the admission of expert testimony than *Frye* was just that, *mere* rhetoric, and that the Court's real agenda was to tighten the standards of admissibility so as to keep out "junk science." And it's true that the Court didn't interpret FED. R. EVID. 702 in the most liberal way it might have done, as requiring only relevance, and only the barest of relevance at that; but took it to require, not only more than bare relevance but, in addition, reliability. All the same, so far as I can tell, the Court's observations to the effect that FED. R. EVID. 702 was more liberal than *Frye* were *not* just a rhetorical smoke-screen.

and some conclusion, decision, etc. That the probability that a match between the defendant's DNA and DNA from the crime scene is random is one in a million, for example, is relevant to whether the defendant is the perpetrator; that many published studies concluded that there is no evidence that Bendectin is teratogenic[77] is relevant to whether the drug can cause birth defects, or caused this plaintiff's child's birth defects specifically—as is the fact that some of those studies didn't distinguish women who took the drug in the period of pregnancy during which the fetal limbs were forming from women who took it later in their pregnancies.[78]

In the terminology of my foundherentist theory of the determinants of the quality of evidence and the degree of warrant of a claim,[79] relevance can be understood by reference to the idea of the *supportiveness* of evidence with respect to a claim. Evidence is *positive* with respect to C if, to some degree, it supports C, i.e., makes it more likely that C is true; it is *negative* with respect to C if, to some degree, it undermines C, i.e., supports not-C; [80] and it is *neutral* with respect to C if it is neither positive nor negative, i.e., makes it neither more, nor less, likely that C is true. Evidence is *positively relevant* to C if it supports C; *negatively relevant* if it supports not-C; and *irrelevant* if it is neither positively nor negatively relevant, i.e., is neutral with respect to C. (This analysis diverges from McCormick's understanding of relevance in two ways: it accommodates both dimensions of relevance, the negative as well as the positive—which is surely a step forward; and, no doubt more controversially, it avoids any identification of degrees of relevance with mathematical probabilities, whether conceived objectively, either as relative frequencies or as propensities, or subjectively, as degrees of belief.).[81]

Unlike relevance, reliability is not a relation, but a property—a property most commonly attributed to procedures, methods, processes, informants, "sources," tools, employees, or friends, but sometimes also

77. As, of course, Merrell Dow pointed out in *Daubert* (and probably all the other Bendectin cases). *See* Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1314 (1995) ("Daubert II").

78. *See, e.g.*, Blum, 33 Phila. Cnty. Rep. at 214–17 (reporting the testimony of Dr. Shapiro).

79. *See* SUSAN HAACK, EVIDENCE AND INQUIRY (2d ed. 2009) (1993); SUSAN HAACK, *Legal Probabilism: An Epistemological Dissent*, EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 122 (2014), *translated in* ESTÁNDARES DE PRUEBA Y PREUBA CIENTÍFICA: ENSAYOS DE EPISTEMOLOGÍA JURÍDICA 65–98 (2013).

80. So in this context I'm using "supportiveness" in a technical, generic sense, to include the negative as well as the positive, the undermining as well as the supporting.

81. *See generally* HAACK, *Legal Probabilism*, *supra* note 79.

used in a slightly oblique sense (as "healthy" is sometimes used of a diet, an exercise regime, or the climate in a certain location) of testimony or a conclusion. A reliable watch or clock is one you can count on to keep time; a reliable car is one you can count on to run well and not break down; a reliable friend is one you can count on to show up when you have arranged to meet, and to provide the help he promised; a reliable plane schedule is one you can count on to tell you when your flight will leave and when it will arrive at its destination; a reliable recipe or a reliable method is one you can count on to give a good result; and a reliable witness is one you can count on both to tell you the truth as he believes it to be, and to know what he's talking about.[82] In short, the reliable is the trustworthy—the source, friend, tool, etc. you can count on.

But what about reliability as it applies to testimony, statements, claims—which is what's most centrally at issue here? The text of *Daubert* suggests that reliable testimony is testimony arrived at by a reliable method; specifically, that reliable expert testimony is testimony arrived at by "scientific" methodology. Justice Blackmun's footnote 9, however, takes a different approach: reliable expert testimony is testimony based on a principle that, as he says, really supports what it purports to.[83] This conforms more closely to what seems to be the natural way to think of reliability as it applies to claims, etc.: a reliable claim is a *well-founded* claim, a claim which is (in my foundherentist terminology) *independently secure*. This is a property of the testimony or the claim, a matter of how well-established it is.[84]

Both Thayer and McCormick sometimes speak of "logical" relevancy;[85] but both implicitly acknowledge that whether and, if so, to what degree, this claim or that alleged fact is relevant to a question at issue usually depends on facts about the world. Thayer writes that in

---

82. W.K. Clifford, *The Ethics of Belief*, THE ETHICS OF BELIEF AND OTHER ESSAYS 70, 77 (Leslie Stephen & Sir Frederick Pollock eds. 1947); *see also*, Susan Haack, *The Expert Witness: Lessons from the U.S. Experience*, 28 HUMANA MENTE: J. PHIL. STUD. 39 (2015); Susan Haack, *Credulity and Circumspection: Epistemological Character and the Ethics of Belief*, PROC. AM. CATH. PHIL. ASS'N (2014).

83. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 590 n.9 (1993) ("Daubert I"). (Of course, Justice Blackmun isn't using "supports" in the technical sense I have given it.).

84. That is, how secure it is, independent of whatever support it may be given by the conclusion it is adduced to support (which is why I call this the "independent security" condition).

85. "[U]nless excluded by some rule or principle of law, all that is logically probative is admissible." THAYER, *supra* note 5, at 265. "[I]f evidence is logically probative it should be received unless there is some distinct ground for refusing to hear it." MCCORMICK, *supra* note 6, at 314.

determining relevancy the law relies on "logic *and general experience*";[86] discussing whether the defendant's having made a suicide attempt is relevant to his consciousness of guilt, McCormick observes that *we have no statistics* to tell us how much this raises the probability that he guilty.[87]

In fact, relevance is rarely a matter of pure logic. Why, for example, is the fact that this character witness is his brother relevant to how credible his testimony is that the defendant wouldn't hurt a fly? Because, by and large, brothers can be expected to have a good opinion of each other's character, and perhaps to be motivated to fudge a bit, if need be, to help each other out. Why is the fact that this forensic technician was disciplined for turning in his paperwork late relevant to how credible his testimony is that the substance seized was cocaine? Because, in his haste to complete his paperwork on time, he may have been sloppy about double-checking the sample numbers. Why is the fact that mice exposed to this substance develop cataracts relevant to whether exposure to the defendant's cleaning fluid caused the plaintiff to do the same? Because mice are physiologically similar to human beings, and the substance to which they were exposed is the main ingredient in the product. Why is the fact that this study was conducted by the manufacturer of the drug[88] relevant to whether its conclusions are likely true? Because we have

86. THAYER, *supra* note 5, at 265 (emphasis added).

87. MCCORMICK, *supra* note 6, at 318–19 (emphasis added).

88. For example, many of the studies of Bendectin referred to earlier were conducted by Merrell Dow, or funded by Merrell Dow; and several were published in *Teratology*, which was edited by medical scientist Robert Brent, a passionate advocate for the drug who was for many years on retainer with Merrell Dow. As of 1996 (the date of Judge Bernstein's devastating summary of flaws in the defendant's expert testimony in *Blum*, 33 Phila. Cnty. Rep. at 230), there were twelve animal studies of Bendectin, and forty epidemiological studies. JOSEPH SANDERS, BENDECTIN ON TRIAL, 66–70 (1998). Five of the animal studies (Hendrickx 1982, Hendrickx 1983, Hendrickx 1985(a) Hendrickx 1985(b), and Tyl 1988), and four of the epidemiological studies (Smithells 1983, Michaelis 1983, Shiono 1989, and Erikson 1991) were published in *Teratology*. *Id.* at 253–275. Five of the animal studies (Gibson 1968, Gibson 1975(a), Gibson 1975(b), Hendrickx 1985, and the second version of Tyl 1988), and six of the epidemiological studies (Bunde and Bowles 1963, Milkovich 1976, Shapiro 1977, Smithells 1978, Mitchell 1981, and Mitchell 1983) were conducted by Merrell Dow staff or funded at least in part by Merrell Dow. *Id.* at 64–65, 69, 97–98, 262, 274. There is also evidence that some co-authors of various of these studies had connections with Merrell Dow. (Not surprisingly, many of these authors and co-authors testified for Merrell Dow in one or more of the Bendectin cases.) A full exploration would, of course, require a whole other paper. *But see generally* Haack, *supra* note 71, at 178-79; Susan Haack, *What's Wrong with Litigation-Driven Science?* 38 SETON HALL L. REV. 1053 (2008), *reprinted in* EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 180 (2014).

reason to believe that studies conducted by manufacturers are more likely to reach conclusions favorable to their products.[89] And so on.

Reliability, too, usually depends on facts about the world. Why are epidemiological studies calling on subjects' own reports of their diseases and disorders highly unreliable? Because people's self-diagnoses are often wrong.[90] Why is this recipe for kheema[91] not perfectly reliable? Because the texture of the resulting dish depends on how juicy the tomatoes are, as well as on how much stock you add. Why is the testimony of a color-blind eye-witness that the car that caused the accident was red unreliable? Because the commonest form of color-blindness is an inability to distinguish red from green. And so forth.

Evidence may be relevant to a given issue, but not reliable; or reliable, but not relevant to the issue. Imagine, for example, that X is on trial for the rape of Y. A jail-house informant's testimony that X confessed to the crime is highly relevant, all right, but very unreliable;[92] a DNA expert's testimony that Y is a descendant of Thomas Jefferson may be highly reliable,[93] but is unlikely to be even remotely relevant to the case. Or imagine that X is suing Y for a disorder allegedly caused by his taking Y's drug. Epidemiological evidence showing that exposure to this drug is associated with a three-fold increase in the risk of developing

---

89. *See, e.g.*, Richard A. Davidson, *Sources of Funding and Outcome of Clinical Trials*, 1 J. GEN. INTERNAL MED. 155 (1986); Lisa A. Bero et al., *The Publication of Sponsored Symposiums in Medical Journals*, 327 NEW ENG. J. MED. 1135 (1992); Paula A. Rochon et al., *A Study of Manufacturer-Supported Trials of Non-Steroidal Anti-Inflammatory Drugs in the Treatment of Arthritis*, 154.2 ANNALS INTERNAL MED. 157 (1994); Lee S. Friedman & Elihu D. Richter, *Relationship between Conflicts of Interest and Research Results*, 19 J. GEN. INTERNAL MED. 51 (2004).

90. The only study that showed an increased risk of connective-tissue disorders among women with silicone breast implants, the Hennekens study, was apparently based on the women's own reports. Charles H. Hennekens et al., *Self-Reported Breast Implants and Connective Tissue Disease in Female Health Professionals: A Retrospective Cohort Study*, 275.8 J. AM. MED. ASS'N 616 (1996).

91. JOSEPH COTTA, A HERITAGE OF INDIAN COOKING 54–55 (1980).

92. "In all, there have been 111 death row exonerations since capital punishment was resumed in the 1970s. The snitch cases account for 45.9% of those. That makes snitches the leading cause of wrongful convictions in U.S. capital cases—followed by erroneous eyewitness identification testimony in 25.2% of the cases, false confessions in 14.4%, and false or misleading scientific evidence in 9.9%." Center on Wrongful Convictions, *The Snitch System: How Snitch Testimony Sent Randy Steidl and Other Innocent Americans to Death Row* (Winter 2004-2005), http://www.innocenceproject.org/causes-wrongful-conviction/SnitchSystemBooklet.pdf; *Informants*, INNOCENCE PROJECT, http://www.innocenceproject.org/causes-wrongful-conviction/informants (last visited Oct. 27, 2015).

93. DNA evidence indicates that *some* male of the Jefferson family was the father of *one* of the children of his house-slave Sally Hemings. WILLIAM G. HYLAND, IN DEFENSE OF THOMAS JEFFERSON: THE SALLY HEMINGS SEX SCANDAL (2009).

this disorder is very relevant, all right; but if the study had no controls, wasn't randomized or double-blinded, relied on the subjects' own reports of their disorders, or, etc., not very reliable. On the other hand, evidence from an impeccably well-designed and well-conducted epidemiological study that this drug triples the risk of an entirely different disorder is likely much more reliable, but probably not relevant, or only marginally so.

My analysis confirms this: evidence may be relevant to a claim but not reliable, i.e., while it supports, or undermines, the claim in question, it lacks justification itself; or reliable but not relevant, i.e., while it is solid enough in itself, it has no bearing either way on the likely truth of that claim.

Most to the present purpose, as such familiar locutions as "highly relevant," "very relevant," "remotely relevant," "marginally relevant," etc., indicate, relevance is normally a matter of degree, depending on *how closely* this fact, that evidence, etc., bears on the matter at hand. That the defendant was found standing over the victim's body holding the murder weapon, for example, is highly relevant to whether he is the perpetrator of the crime; that one of the dozen scientists who conducted the rabbit tests on a drug had been sleeping poorly since his recent divorce, if it is relevant at all to whether the drug is toxic to humans, is only marginally so.

And, as such familiar locutions as "perfectly reliable," "highly reliable," "very reliable," "fairly reliable," "somewhat unreliable," "not very reliable," etc., indicate, reliability, like relevance, is also a matter of degree. My watch is reasonably reliable, keeping time more than adequately for my day-to-day purposes; but it has been known to gain a minute or two, and isn't what I'd use if my life depended on my being *exactly* on time. This recipe for kheema is very reliable, but not *perfectly* so—when the tomatoes you use are exceptionally large and juicy, you should add less stock than it says. This publication is usually a reliable source with respect to the various régimes governing the admissibility of expert testimony across the U.S., but not *entirely* so, since in 2014 the account in the text was inconsistent with the information given in one of the notes.[94] That website is nothing but *un*reliable, self-serving hype. Epidemiological studies calling on subjects' own reports of the disorders they suffer are *less* reliable than studies that require medical diagnoses; double-blind, controlled, randomized clinical trials are the *most* reliable,

94. DEMOSTHENES LORANDOS & TERENCE CAMPBELL, CROSS EXAMINING EXPERTS IN THE BEHAVIORAL SCIENCES §1:16:1 (2014). According to their notes 5 and 6, thirty-six states are now "*Daubert* or *Daubert*-leaning," and twelve continue to use *Frye*; but in the text the authors seem to have miscounted!

the "gold standard" of epidemiology. The testimony of this eye-witness, who has perfect vision, is *more* reliable than the testimony of that eye-witness, who is short-sighted and color-blind. DNA-identification techniques are *more* reliable than fingerprint-identification techniques. And so on.

This, too, is confirmed by my epistemological analysis, which is gradational through and through. Evidence is *more* relevant to conclusion C, the *more* it tends to support, or to undermine, C: highly relevant if it supports C (or supports not-C) to a high degree, marginally relevant if it gives only very modest support to C (or to not-C). And it is *more* reliable the *more* secure it is.

3. INTERPRETIVE DISAGREEMENTS

Even before *Daubert*, some courts were (understandably) confused about the relation between relevance and reliability. For example, in *Downing* Judge Becker had hinted that reliability might be something like a degree of relevance, writing that "'[h]elpfulness' necessarily implies a quantum of reliability beyond that required to meet a standard of bare logical relevance."[95] And the year before *Daubert*, an appeals court in Texas had written that "before novel scientific evidence may be admitted under [Texas] Rule 702, the proponent must persuade the trial court, by clear and convincing evidence, that the evidence is reliable and *therefore* relevant."[96] But I want to focus here on the most consequential conceptual complication—the gradational character of both relevance and reliability, and the resulting mismatch with the categorical concept of admissibility—and how this plays out in legal decisions applying the new evidentiary régime for expert testimony put in place by *Daubert*.

Evidence E may be highly relevant to claim C, or somewhat relevant, or only marginally relevant; and it may be highly reliable, fairly reliable, or quite shaky. But an expert's testimony must be ruled *either* admissible, *or else* inadmissible.[97] Briefly and roughly: if the degree of

95. United States v. Downing, 753 F.2d 1224, 1235 (1985).

96. Kelly v. State, 824 S.W.2d 568, 573 (Tex. Crim. App. 1992) (emphasis added). The Texas Rules of Evidence closely mirrored the FRE. (This observation, though awkwardly put, raises a real question, which, however, I shall have to set aside here: should we count as *actually* relevant testimony that *would be* relevant if it were well-founded, even if it is known to be highly unreliable?)

97. A point first made in print, so far as I know, by Dale Nance in *Two Concepts of Reliability*. Dale Nance, *Two Concepts of Reliability*, AM. PHIL. ASS'N NEWS. PIL. & L., Fall 2003, at 123. (However, as his title indicates, Nance was writing only of reliability.) It is worthy of note that, proposing that England and Wales adopt a *Daubert*-style rule for expert testimony, the Law Commission inadvertently highlighted the mismatch when it urged that such testimony should be admissible only if it is "*sufficiently* reliable." Law

relevance and reliability required is minimal, *Daubert* gatekeeping will be, as Justice Blackmun apparently anticipated, flexible and accommodating; but the higher the degree of relevance or reliability is required, the less flexible and accommodating it will be, and the likelier it is that decisions about the admissibility of expert testimony will shade into determinations of its sufficiency.

The year after *Daubert*, in *In re Paoli Railroad Yard*,[98] we find Judge Becker wrestling with precisely these interpretive issues. He begins by asking how high a threshold of reliability is required by *Daubert*;[99] and argues that, while the standard must be something more than a prima facie showing,[100] it should not be *so* high that the proponent of the evidence is in effect asked to prove his case twice, first at the admissibility stage, and then again at trial:[101]

> [The reliability prong of *Daubert*] does not mean that plaintiffs have to prove their case twice—they do not have to demonstrate to the judge by a preponderance of the evidence that the assessments of their experts are *correct*, they only have to demonstrate by a preponderance of evidence that their opinions are reliable.[102]

He continues:

> The grounds for the expert's opinion merely have to be good, they do not have to be perfect. The judge might think that there are good grounds for an expert's conclusion even if the judge thinks that there are better grounds for some alternative conclusion . . .[103]

Commission, *Expert Evidence in Criminal Proceedings in England and Wales*, (Law Comm. No. 325, 2011) (emphasis added). (It is almost, but perhaps not quite, too obvious to need saying that adding "sufficiently" provides exactly nothing by way of substantive guidance.) The government subsequently declined to accept the Law Commission's proposal. Ministry of Justice, *The Government's Response to the Law Commission Report, "Expert Evidence in Criminal Proceedings in England and Wales,"* (Law Com. No. 325) (2013); *In re* Paoli R.R. Yard Litig., 35 F.3d 717 (1994).

98. *Paoli*, 35 F.3d at 717.

99. *Id.* at 743.

100. *Id.*

101. *Id.* at 744.

102. *Id.* (emphasis in original).

103. *Id.*

The reliability requirement, he writes, "must not be used as a tool by which the court excludes all questionably reliable evidence."[104] Similarly, he continues, for relevancy: the standard is higher than "bare" relevance, but not *very* high.[105] In short:

> [T]he primary limitation on the judge's admissibility determinations is that the judge should not exclude evidence simply because he or she thinks that there is a flaw in the expert's investigative process which renders the expert's conclusions incorrect. The judge should only exclude the evidence if the flaw is large enough that the expert lacks "good grounds" for his or her conclusions.[106]

But in some other post-*Daubert* cases we encounter much more stringent interpretations both of relevance and of reliability than Judge Becker envisages. Indeed, a significantly higher degree of relevance seems to be required in Judge Kozinski's final ruling in *Daubert* on remand from the Supreme Court;[107] and a significantly higher degree of reliability seems to be suggested, only a couple of years later, in the Supreme Court's ruling in *Joiner*.[108]

i. Relevance and Sufficiency: The Legacy of Daubert II.

Judge Kozinski's final ruling in *Daubert* is a rhetorical *tour de force*; it is also, however, more than a little confusing (not to say more than a little confused). For example, pointing out that the plaintiffs' experts' work on Bendectin had all been conducted specifically for the purposes of litigation, Judge Kozinski proposes what is sometimes referred to as a fifth *Daubert* factor—whether the science is litigation-driven—to add to the Supreme Court's flexible list of indicia of evidentiary reliability.[109] But when he rules that the Dauberts' experts' testimony would have to be excluded under the new *Daubert* régime, as it had previously been under *Frye*, he doesn't rely on this new *Daubert* factor, but argues that the testimony of all but one of these experts is inadmissible on grounds of irrelevance;[110] and when he finds that the only expert whose testimony

104. *Id.* (quoting *In re* Paoli R.R. Yard PCB Litig. 916 F.2d 829, 857 (1990)).

105. *Id.* at 745.

106. *Id. at* 746.

107. Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311 (1995) ("Daubert II").

108. Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).

109. *Daubert II*, 43 F.3d at 1317; *see also* Haack, *What's Wrong with Litigation-Driven Science?*, *supra* note 88.

110. *Daubert II*, 43 F.3d. at 1320–21.

would satisfy the "fit" requirement, Dr. Palmer, flunks the reliability requirement, it is on the grounds, *not* that his work is litigation-driven, but that he "offers no tested or testable theory" to explain how he was able to rule out other possible causes of Jason Daubert's birth defects.[111]

On the question of reliability, Judge Kozinski apparently agrees with Judge Becker; at any rate, he observes that the fact that scientific work has been scrutinized through the peer-review process shows that it at least meets "minimal" scientific standards.[112] But what concerns me here is Judge Kozinski's interpretation of the relevance requirement. The Supreme Court's reference to "fit," he argues, indicates that something more is required than "bare" (i.e., presumably, marginal or remote) relevance.[113] And what *is* that something more? Well, he writes, "California tort law requires plaintiffs to show . . . that Bendectin . . . more likely than not caused their injuries," and:

> In terms of statistical proof, this means that plaintiffs must establish not just that their mothers' ingestion of Bendectin increased somewhat the likelihood of birth defects, but that it more than doubled it—only then can it be said that Bendectin is more likely than not the source of their injury.[114]

But "none of plaintiffs'. . .experts [except Dr. Palmer] claims that ingestion of Bendectin during pregnancy more than doubles the risk of birth defects";[115] and so they flunk the second prong of *Daubert*, which "goes primarily to relevancy."[116] This extraordinary argument (a) equates degrees of proof with statistical probabilities; (b) assesses each expert's testimony individually; and (c), most to the present purpose, raises the standard of *admissibility* under the relevance prong to the standard of *proof.*

As I have argued elsewhere, degrees of proof are best construed as degrees of warrant of the proposition at issue by the evidence presented; and these degrees of warrant *can't* be identified with statistical probabilities.[117] You can't infer, from the *statistical probability* that the risk of developing disorder D among those exposed to substance S is more than double the risk among those not so exposed, to the conclusion

---

111. *Id.* at 1319, 1321–22.
112. *Id.* at 1318.
113. *Id.* at 1321 & n.17.
114. *Id.* at 1320
115. *Id.* at 1320–21.
116. *Daubert II*, 43 F.3d at 1321 & n.17 (citing Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 591 (1993) ("Daubert I").
117. *See* HAACK, *Legal Probabilism*, *supra* note 79, at 56–64.

that it is more epistemologically likely than not that this plaintiff's D was caused by his exposure to S[118]—any more than you can infer, from the statistical probability that the match between DNA from the crime scene and DNA from the defendant is random is one in a million, to the conclusion that the epistemic likelihood that the defendant is not guilty is, also, a one in a million.[119]

Moreover, as I have also argued elsewhere, a piece of evidence that, by itself, seems only marginally relevant to a conclusion may, in the context of other evidence, turn out to be highly relevant—but may be excluded by a too-atomistic screening process.[120] "A brick is not a wall," Dean McCormick famously observed;[121] but Judge Kozinski's atomistic approach could exclude the bricks a party needs to build its wall simply on the grounds that they *are* just bricks!

But it's the third aspect of Judge Kozinski's argument that's most to the present purpose. In effect, what he's asking is that each plaintiff's expert produce evidence which (assuming it reliable) would be sufficient to reach the standard of proof. As Prof. Imwinkelried would observe shortly after *Daubert II*, with this argument Judge Kozinski "breathed new life"[122] into the relevance prong—new life that seemed to elide admissibility into sufficiency.

Appeal to the idea of more than doubled relative risk [RR > 2] wasn't new: Judge Kozinski cited *DeLuca v. Merrell Dow* (1990);[123] well before *DeLuca* the idea had played a role in some vaccine cases;[124] and in 1994 it had been endorsed in the first edition of the new *Reference Manual on Scientific Evidence*, where the authors of the chapter on epidemiology wrote that "[t]he threshold for concluding that an agent was more likely the cause of a disease than not is a relative risk greater

118. SUSAN HAACK, *Risky Business: Statistical Proof of Specific Causation*, EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 264 (2014).

119. *See* William C. Thompson & E. L. Schuman, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167 (1987).

120. *See* Susan Haack, *Proving Causation: The Weight of Combined Evidence*, 4 J. HEALTH & MED. L. 253 (2008), *reprinted in* EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 208 (2014).

121. MCCORMICK, *supra* note 6, at 317.

122. Edward J. Imwinkelried, *The Second Prong of the Daubert Test: Disturbing Implications of Two Recent Civil Cases*, 33 CRIM. L. BULL. 570, 572 (1997).

123. DeLuca v. Merrell Dow Pharm., Inc., 911 F.2d 941, 958 (3rd Cir. 1990).

124. Cook v. United States, 545 F. Supp. 306 (N.D. Cal. 1982); Manko v. United States, 636 F. Supp. 1419 (W.D. Mo. 1986). The story is told in some detail in HAACK, *Risky Business: Statistical Proof of Individual Causation*, *supra* note 118, at 269–74.

than 2.0."[125] What *was* new was the idea that, in toxic-tort cases, this set the standard for admissibility under *Daubert*—an idea that soon began to spread as many other courts cited Judge Kozinski's argument and adopted the same lofty standard of relevance in their admissibility decisions: for example, *Ambrosisi v. Upjohn* (1995);[126] *Sanderson v. International Flavors and Fragrances* (1996);[127] *Hall v. Baxter Healthcare* (1996);[128] *Schudel v. G.E* (1997);[129] *Bartley v. Euclid* (1998);[130] *In re Hanford Nuclear Reservation* (1998);[131] *In re Breast Implant Litigation* (1998);[132] and *Allison v. McGahn* (1999).[133]

The second edition of the *Reference Manual*, published in 2000, included a new chapter on epidemiology, under a new lead author; but the section on the RR > 2 idea is much like the one in the first edition—only now including a number of citations to some of the cases mentioned

125. Linda A. Bailey, Leon Gordis & Michael D. Green, *Reference Guide on Epidemiology*, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 123, 168 (Federal Judicial Center, 1st ed. 1994).

126. *See, e.g.*, Ambrosini v. Upjohn Co., [1996 Transfer Binder] Prod. Liab. Rep. (CCH) ¶ 14,462, 47,144 (D.D.C. Oct. 26, 1995) (ruling that Dr. Strom's testimony was inadmissible because he wasn't able to say that a mother's having taken Depo-Provera more than doubled the risk of birth defects in her baby). This decision, however, was overturned the next year, the appeals court ruling that Dr. Strom's testimony "comfortably cleared the hurdle of admissibility established by Daubert." Ambrosini v. LaBarraque, 101 F.3d 129, 140 (D.D.C. 1996).

127. Sanderson v. Int'l Flavors & Fragrances, Inc., 950 F. Supp. 981, 1000 (C.D. Cal. 1996) (ruling that none of Ms. Sanderson's experts were admissible because they weren't able to testify that exposure to the defendants' products more than doubled her risk of developing chemical sensitivities).

128. Hall v. Baxter Healthcare Corp., 947 F. Supp. 1387, 1398, 1403 (D. Or. 1996) (finding that "[i]n epidemiological terms, Oregon's standard of proof means that plaintiffs must be able to show a relative risk of greater than 2.0"; and "[u]nder this substantive standard, if an expert cannot state the causal connection in terms of probability or certainty, [their] testimony must be excluded under the [relevance] prong of Rule 702").

129. Schudel v. Gen. Elec. Co., 120 F.3d 991, 996 (9th Cir. 1997) (adopting Judge Kozinski's statistical gloss on "preponderance of the evidence").

130. Bartley v. Euclid, 158 F.3d 261, 273 (5th Cir. 1998) ("[a]ssuming, without deciding, that *Havner*'s rule controls. . ."), *vacated by* 169 F.3d 215 (5th Cir. 1999). *See* Havner, *infra* note 144 and accompanying text.

131. *In re* Hanford Nuclear Reservation Litig., No. CY091-3015-AAM, 1998 WL 775340, at *8 (E.D. Wash. Aug. 21, 1998) (accepting Judge Kozinski's statistical interpretation of the standard for allowing the jury to hear plaintiffs' causation evidence), *rev'd*, 292 F.3d 1124 (9th Cir. 2002).

132. *In re* Breast Implant Litig., 11 F. Supp. 2d 1217 (D. Colo. 1998) (finding that, for their evidence of medical causation to be relevant, plaintiffs must show a more than doubled risk).

133. Allison v. McGahn Med. Corp., 184 F.3d 1300 (11th Cir. 1999) (citing FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 168–69 (1st ed. 1994).

here.[134] Naturally, federal courts continued to follow Judge Kozinski's lead: in *In re Silicone Gel Implants Products* (2004),[135] for example, and *Henricksen v. Conoco Phillips* (2009).[136] The chapter on epidemiology in the third edition of the *Reference Manual*, published in 2011, finally acknowledged that the equation of the preponderance standard with a showing of RR > 2 required a number of caveats.[137] But courts haven't always paid attention to these (much-needed) notes of caution.

In *Samaan v. St. Joseph Hospital* (2012),[138] for example, affirming the trial court's exclusion of Dr. Tikoo's testimony that, had the plaintiff been given a timely injection of t-PA,[139] he likely would not have suffered stroke-related injuries, the court of appeals argues that this testimony doesn't reach the required standard of relevance because it fails to show that, had he been given the injection, Mr. Samaan would have had a greater than 50% chance of the better outcome.[140] It is simply taken for granted that the relevancy prong requires a showing of RR > 2, and that this is equivalent to the preponderance standard. And in a California case, *Cooper v. Takeda Pharmaceuticals* (2015),[141] ruling that the trial court had abused its discretion in excluding the testimony of the Coopers' expert urologic oncologist, the court of appeals avers, citing *Daubert II* and the second edition of the *Reference Manual*, that the expert's testimony *did* meet the relevancy standard—it established that

134. Michael D. Green, D. Michal Freedman & Leon Gordis, *Reference Guide on Epidemiology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 333, 384 (Federal Judicial Center, 2d ed. 2000).

135. *In re* Silicone Gel Implants Prod. Liab. Litig., 318 F. Supp. 2d 879, 893 (C.D. Cal. 2004) (averring that "under California law [statistical] analyses must show a relative risk greater than 2.0 to be 'useful' to the jury").

136. Henricksen v. Conoco Phillips Co., 605 F. Supp. 2d 1142, 1158 (E.D. Wash. 2009) (excluding the plaintiff's expert testimony on the grounds that in the Ninth Circuit an epidemiological study can be probative of specific causation "only if [it] shows the relative risk is greater than 2.0").

137. Michael D. Green, D. Michal Freedman & Leon Gordis, *Reference Guide on Epidemiology*, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 612–18 (3d ed. 2011). I notice that these caveats bear quite a close resemblance to some points I had made in *Risky Business*. HAACK, *supra* note 118, at 285–290, which in 2011 was in press, and circulating, but not yet published.

138. Samaan v. St. Joseph Hosp., 670 F.3d 21 (1st Cir. 2012).

139. Tissue Plasminogen Activator for Acute Ischemic Stroke, which improves the outcome for some stroke patients if administered within three hours of the onset of stroke symptoms. Samaan v. St. Joseph Hosp., 755 F. Supp. 2d 236, 239 (D. Me. 2010).

140. *Samaan*, 670 F.3d at 33.

141. Cooper v. Takeda Pharm. Am., Inc., 239 Cal. App. 4th 555 (2015) *rev'g* Coordination Proceeding Spec. Title (Rule 2.550) Actos Prod. Liab. Cases, No. CGC-12-518535, 2013 WL 1846220 (Sup. Ct. Cal. 2015).

Takeda's diabetes drug Actos was more likely than not the cause of Mr. Cooper's bladder cancer.[142]

Judge Kozinski's understanding of the relevancy requirement of *Daubert* also made its way into state courts. For example, it seemed to be endorsed by the Supreme Court of Texas in *Merrell Dow Pharmaceuticals v. Havner* (1997),[143] and by the Supreme Court of Vermont in *Estate of George* (2010).[144]

But other courts, and some dissenting judges, took a different line. Some (rightly) resisted the equation of statistical probabilities with degrees of proof: for example, the courts in *Jones v. Owens-Corning* (1996);[145] *Pick v. American Medical Systems* (1997);[146] *McDaniel v. CSX Transportation* (1997);[147] *Minnesota Mining and Manufacturing v. Atterbury* (1998);[148] and the court of appeals that overturned the earlier ruling in *In re Hanford Nuclear Reservation* (2002).[149] More to the present purpose, the court in *Pick*, noting that FRE 401 defines relevant evidence as evidence having "'*any* tendency' to prove or disprove a fact of consequence in the case," argued that a showing of more than doubled risk is *not* required for epidemiological evidence to be admissible.[150] And, most to the present purpose, in a passage reminiscent of Judge Becker's argument in *Paoli*, Vermont Chief Justice Reiber observed in his dissent in *Estate of George* that to impose a requirement that, to be even admissible, epidemiological evidence must show a more than doubled relative risk, "sets a threshold that requires each study to *prove* that claimant[s] should win on the merits";[151] and objects that this is

142. *Cooper*, *supra* note 141, at 593–94.

143. Merrell Dow Pharm., Inc., v. Havner, 953 S.W.2d 706, 717 (Tex. 1997) (averring that "there is a rational basis for relating the requirement that there be more than a 'doubling of the risk' to our. . .evidence standard of review and to the more likely than not burden of proof").

144. Estate of George v. Vermont League of Cities and Towns, 993 A.2d 367, 375 (Vt. 2010) (finding that the trial court had not erred in taking a relative risk greater than 2 as a benchmark for admissibility of plaintiff's experts, given that this "easily tied into Vermont's 'more likely than not' civil standard").

145. Jones v. Owens-Corning Fiberglass Corp., 672 A.2d 230, 234–35 (N.J. Super. Ct. App. Div. 1996) (rejecting the more-than-doubled-risk threshold).

146. Pick v. Am. Med. Sys., Inc., 958 F. Supp. 1151, 1160 (E.D. La. 1997).

147. McDaniel v. CSX Transp., Inc., 955 S.W.2d 257, 264 (Tenn. 1997) (rejecting the idea that the more-than-doubled-risk test be adopted as a matter of law).

148. Minnesota Mining & Mfg. Co. v. Atterbury, 978 S.W.2d 183, 198 (Tex. App. 1998) (finding that there is no requirement of a showing of more than doubled risk in toxic tort cases).

149. *In re* Hanford Nuclear Reservation Litig., 292 F.3d 1124, 1135–37 (9th Cir. 2002) (finding that the lower court had erred in requiring a showing of more than doubled risk).

150. *Pick*, 958 F. Supp. at 1160.

151. *Estate of George*, 993 A.2d at 387 (Reiber, J. dissenting) (emphasis in original).

inconsistent with the accepted principle that "admitted evidence does not alone have to meet the proponent's burden of proof on a particular issue."[152]

ii. Reliability and Sufficiency: The Legacy of Joiner.

One could hardly describe Chief Justice Rehnquist's ruling for the Supreme Court in *Joiner*,[153] like Judge Kozinski's final ruling in *Daubert*, as a rhetorical *tour de force*. But, like Judge Kozinski's, Justice Rehnquist's ruling pushes admissibility closer to sufficiency—this time, however, not by raising the degree of relevance, but by potentially raising the degree of reliability required. Justice Rehnquist doesn't, however, do this explicitly: rather, it's the result of his strategically sidestepping some key issues.

The Supreme Court took *Joiner*, the second case in its "*Daubert* trilogy," to settle the question of the standard of appellate review for evidentiary decisions under *Daubert*. The district court had excluded Mr. Joiner's proffered expert testimony that his occupational exposure to PCBs [polychlorinated biphenyls] had promoted his early development of small-cell lung cancer, and—given that, with Joiner's experts excluded, there was no case to answer—granted summary judgment to General Electric [G.E.]. The court of appeals reversed, arguing that, since FRE 702 displays a preference for admissibility, in cases such as this courts should apply a "particularly stringent standard of review."[154] At the Supreme Court, G.E.'s attorneys argued that this amounted to a new, and improper, standard of review; Mr. Joiner's attorneys replied that, on the contrary, it had simply applied the regular "abuse of discretion" standard with the level of rigor necessary in such instances.[155] Sidestepping this argument, Justice Rehnquist wrote on behalf of the Supreme Court that *Daubert* hadn't changed the standard of review of evidentiary exclusions, which remained the same: abuse of discretion.[156] Moreover, he continued, in this instance the trial court had *not* abused its discretion in excluding Joiner's experts.[157]

However, Mr. Joiner's attorneys had also argued that their experts used the very *same* "weight of evidence" methodology as the experts for the defendant, G.E.; so that, given *Daubert*'s insistence that the focus of

152. *Id.*
153. Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).
154. *Id.* at 140 (citing Joiner v. Gen. Elec. Co., 78 F.3d 524, 529 (11th Cir. 1996)).
155. *Id.* at 141.
156. *Id.*at 142.
157. *Id.* at 146–47.

a reliability inquiry should be on an expert's methodology, not his conclusions, legally it had been an error to exclude their experts while admitting G.E.'s. But Justice Rehnquist sidesteps this argument too, first announcing that there is, after all, no real distinction between methodology and conclusions,[158] and then urging that what courts should ask in making reliability determinations isn't whether the expert used proper methodology, but whether there is "too great an analytical gap between [his] data and the opinion proffered."[159] He points out that "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert."[160] Granted. However, he tells us nothing about what, exactly, an "analytical gap" is, or how a court is to determine how large a gap is *too* large for expert evidence to be admissible. If the conclusion rests on nothing but the expert's say-so, it's inadmissible, yes; but when, *short of that*, is there an "analytical gap" too large to be acceptable?

Thus far, I have presented *Joiner* as suggesting a stronger interpretation of reliability. However, an "analytical gap" is presumably a lacuna between data and a conclusion—which, being a relation, looks more like relevance than reliability; so it seems as if *Joiner* is also blurring the distinction between the two distinct requirements, reliability and relevance, that the *Daubert* Court found implicit in the "helpfulness" clause of Rule 702. Still, the "analytical gap" terminology is explicitly offered as articulating reliability; and this, with the extension of the scope of judicial gatekeeping from experts' methodology to their conclusions, increased the likelihood that a higher degree of reliability might be required than before; and so, like the ruling in *Daubert II*, more likely that admissibility might shift closer to sufficiency.

The Supreme Court's ruling in *Kumho Tire*,[161] two years later, arguably represents another incremental step in the same direction. *Daubert* applies, the *Kumho* Court ruled, to all expert testimony, not just

158. *Id.* at 146. As Justice Stevens pointed out in his partial dissent, it is doubtful whether this element in the ruling in *Joiner* can really be reconciled with *Daubert*, in which the distinction between methodology and conclusions played a starring role. *Id.* at 152 (Stevens, J. dissenting in part); *see also In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 741-52 (3rd Cir. 1994).

159. *Joiner*, 522 U.S. at 146. The "analytical gap" terminology wasn't new; it is found, for example, in *Turpin*, where the court writes: "The analytical gap between the evidence presented and the inferences to be drawn on the ultimate issue of human birth defects is too wide. Under such circumstances, a jury should not be asked to speculate on the issue of causation." Turpin v. Merrell Dow Pharm., Inc., 959 F.2d 1349, 1360–61 (6th Cir. 1992).

160. Joiner, 522 U.S. at 146.

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

the scientific; but those *Daubert* factors—indicia of reliability crafted for a case where the expert testimony at issue was epidemiological, toxicological, etc.—may not be appropriate where other, non-scientific kinds of expertise are concerned.[162] And the language of the 2000 revision of FRE 702, requiring *inter alia* that expert testimony be based on "sufficient" facts or data,[163] may also have played some role.

Naturally, the "analytical gap" terminology soon began to be heard in decisions on exclusions under *Daubert*. In *Wills v. Amerada Hess Corp.* (2002), for example, affirming the district court's exclusion of plaintiff's expert Dr. Bidanset, the court writes that this expert witness "is using a controversial theory that some toxins do not follow the dose-response relationship, but that any amount of exposure causes cancer."[164] And "[e]ven though benzene and PAHs [polycyclic aromatic hydrocarbons] have been shown to cause some types of cancer, it is *too difficult a leap* to allow testimony that says any amount of exposure to these toxins caused squamous cell carcinoma of the head and neck in the Decedent."[165] Two years later, in *Burleson v. Texas Department of Criminal Justice*, affirming the district court's exclusion of the plaintiff's expert testimony, citing the "analytical gap" language of *Joiner*, and describing the expert's proffered testimony as based on "speculation, guesswork, and conjecture," the court concludes that this was, indeed, nothing more than the expert's *ipse dixit*.[166] The same goes for *Knight v. Kirby Inland Marine* (2005), where, affirming the exclusion of the plaintiff's expert testimony, and screening each expert's testimony one by one, the court cites the "*ipse dixit*" sentence of *Joiner* and observes that it "can, and does, 'conclude that there is simply too great an analytical gap between the data and the opinion proffered.'"[167]

---

162. *Id.* at 150.

163. FRE 702 now reads (after the 2011 "restyling"): a qualified witness "may testify in the form of an opinion or otherwise if (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on *sufficient* facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case." FED. R. EVID. 702 (emphasis added).

164. Wills v. Amerada Hess Corp., No. 98 CIV. 7126(RPP), 2002 WL 140542, at *15 (S.D.NY. Jan.31, 2002).

165. *Id.* (emphasis added).

166. Burleson v. Tex. Dep't of Criminal Justice, 393 F.3d 577, 587 (5th Cir. 2004). It is worth observing that here the difficulties in understanding what an analytical gap is are compounded by the difficulties involved in determining when, exactly, what an expert offers is nothing more than his *ipse dixit*.

167. Knight v. Kirby Inland Marine, 363 F. Supp. 2d 859, 867 (N.D. Miss. 2005).

The "analytical gap" language of *Joiner* also migrated to some *Frye* states. In *Blackwell v. Wyeth* (2009),[168] for example, a Maryland case where it was alleged that a vaccine caused a child's autism, we read, "Generally accepted methodology . . . must be coupled with generally accepted analysis in order to avoid the pitfalls of an 'analytical gap.' Dr. Geier's faulty extrapolation from VAERS [Vaccine Adverse Effect Reporting System] data, a potentially reliable source, manifests the *ipse dixit* identified in the Joiner opinion."[169] And Justice Battaglia notes cases from two other *Frye* states where the "analytical gap" terminology had also been used: *Goeb v. Tharaldson* (Minnesota, 2000),[170] and *Kane v. Motorola, Inc.* (Illinois, 2002).[171]

Professors Green and Sanders claim that, especially since *Joiner* put the emphasis on analytical gaps, there has been a trend towards courts' taking reliability to require that the challenged evidence be sufficient to meet the standard of proof.[172] The evidence they offer falls well short of showing this; but what *is* true is that at least a few courts seem to have gone part of the way, and some, arguably, all the way, towards equating reliability and sufficiency.

I start with *Glastetter v. Novartis Pharmaceuticals* (2000).[173] A mother who suffered an intracerebral hemorrhage [ICH] after taking the anti-lactation drug Parlodel sued the manufacturer, alleging that the drug was the cause. Novartis moved to exclude the plaintiff's expert testimony and for summary judgment. The district court's ruling granting both motions begins with a discussion of the standards for summary judgment, noting specifically that "[t]he trial court may not consider the credibility of the witnesses or weigh the evidence";[174] and continues with a statement of the *Daubert-Joiner-Kumho Tire* standard of admissibility.[175] And then Judge Webber deploys quite a subtle argument: that, while differential diagnosis is a legitimate scientific method, it can only yield conclusions about *specific* causation, not about *general* causation; so that the plaintiff's experts can rely on it to establish that Parlodel caused Mrs. Glastetter's ICH only if there are already

168. Blackwell v. Wyeth, 971 A.2d 235 (Md. 2009).

169. *Id.* at 255.

170. Goeb v. Tharaldson, 615 N.W.2d 800, 816 (Minn. 2000).

171. Kane v. Motorola, Inc., 779 N.E.2d 302, 309 (Ill. 2002).

172. Michael D. Green & Joseph Sanders, *Admissibility Versus Sufficiency: Controlling the Quality of Scientific Testimony in the United States*, UNCERTAIN CAUSATION IN TORT LAW 203 (Miquel Martin-Casals & Diego M. Papayannis eds., 2016).

173. Glastetter v. Novartis Pharm. Corp., 107 F. Supp. 2d 1015 (E.D. Mo. 2000).

174. *Id. at* 1017 (citing White v. Pence, 961 F.2d 776, 779 (8th Cir.1992)).

175. *Id.* at 1018.

grounds for believing that the drug *can* cause such injuries[176]—these experts' differential diagnoses ruling *out* other possible causes of her injury can't show that Parlodel must have been the cause unless the drug has already been ruled *in*, i.e. shown to be a possible cause.

But, Judge Webber continues, plaintiffs have failed to establish this general causation claim. "[T]he case reports, including the re-challenge/de-challenge studies, are *not sufficient to establish* the reliability of plaintiff's experts' causation opinions";[177] "[o]verall, Dr. Kulig's testimony as to causation of vasospasm by Parlodel in humans is *inconclusive*";[178] "the court does not find that. . .sources [referred to by plaintiffs' experts] *establish the reliability* of [their] testimony on whether Parlodel could cause the ICH at issue in this case";[179] the fact that the FDA [Food and Drug Administration] had warned that the risk that bromocriptine mesylate (the active ingredient in Parlodel) may cause a serious adverse effect is unacceptable "*do[es] not establish* that Parlodel caused Mrs. Glastetter's ICH";[180] allegedly hidden company documents related to their animal studies "*fail to establish the reliability of plaintiffs' experts' conclusions.*"[181] Shifting up and back between "plaintiff's experts fail to establish the *reliability* of the proposition that Parlodel can cause ICH" and "plaintiff's experts fail to establish *the proposition* that Parlodel can cause ICH," Judge Webber seems to assume that no item of expert testimony is admissible unless it is, by itself, sufficient to establish a disputed fact.

The argument of *Glastetter* is atomistic; but in some other cases where the line between reliability and sufficiency is blurred the reasoning is, on the contrary, holistic—*so* holistic that it requires plaintiffs trying to show that their own testimony meets the reliability requirement to show what's wrong with the defendants' contrary expert testimony. In *In re Phenylpropanolamine* [PPA] *Products Liability* (2003),[182] for example—multi-district litigation alleging that PPA, widely used in over-the-counter and prescription cough and cold medicine and appetite suppressants, caused hemorrhagic and ischemic stroke in men and women, adults and children, as well as seizures, psychoses, and cardiac

176. *Id.* at 1027 (citing Hall v. Baxter Healthcare Corp., 947 F. Supp. 1387, 1413 (D. Or. 1996)).

177. *Id.* at 1031 (emphasis added).

178. *Id.* at 1032 (emphasis added).

179. *Id.* at 1033 (emphasis added).

180. *Id.* at 1035 (emphasis added). In this context the court observes that the FDA is using a lower standard than tort law. *Id.* at 1036.

181. *Id.* (emphasis added).

182. *In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230 (W.D. Wash. 2003).

and other injuries—defendants moved to exclude plaintiffs' expert testimony under FRE 702 and *Daubert*. Denying this motion in part (with respect to plaintiff's expert testimony regarding strokes), and granting it in part (with respect to their testimony regarding other disorders and any injuries occurring more than three days after exposure to PPA), Judge Rothstein cites Judge Kozinski's strictures about litigation-driven science[183] and the "analytical gap" terminology of *Joiner*.[184] She tells us that she "reviewed pleadings filed in support of and in opposition to the motion, along with the remainder of the record, and . . . heard oral argument and expert testimony. . . ."[185] In short, she conducted a thorough scrutiny of the evidence and arguments of both parties. As a result, while with respect to the part of the defendants' motion that she denies Judge Rothstein acknowledges the distinction between admissibility and weight,[186] the ruling as a whole seems to bring the two closer together.

A more recent example is *In re Zoloft* (2014),[187] multi-district litigation alleging that, when taken during pregnancy, this anti-depression medication caused birth defects. Here too the defendant company moved to exclude plaintiffs' expert testimony. Granting the motion in part (with respect to testimony that Zoloft could cause birth defects), but denying it in part (with respect to testimony that there is a plausible biological mechanism by which it could alter embryonic development),[188] Judge Rufe argues that reliable testimony about human causation should generally be supported by epidemiological studies, and that "when epidemiological studies are equivocal or inconsistent with a causation opinion, experts asserting causation opinions *must thoroughly analyze the strengths and weaknesses of the epidemiological research and explain why [it] does not contradict or undermine their opinion.*"[189] Like Judge Rothstein in *In re PPA*, with respect to the part of the plaintiffs' expert testimony he deems admissible, Judge Rufe acknowledges the difference between admissibility and sufficiency;[190] but, when it comes to the part of their testimony he deems *in*admissible, his argument seems to be that, *in light of the defendant's epidemiological evidence, the plaintiffs' expert testimony is insufficient.*

183. *Id.* at 1238.
184. *Id.*
185. *Id.* at 1234.
186. *Id.* at 1240.
187. *In re* Zoloft Prods. Liab. Litig., 26 F. Supp. 3d 466 (E.D. Pa. 2014).
188. *Id.* at 481–82.
189. *Id.* at 475 (emphasis added).
190. *Id.* at 481.

Perhaps the clearest example is a silicone breast-implant case, *Norris v. Baxter Healthcare*,[191] two years after *In re PPA*. The district court had concluded that the plaintiff's experts' methodology couldn't be scientifically valid because it ignored the many epidemiological studies that found no link between silicone breast implants and systemic disease. Affirming on appeal, observing that he can't find a single case in which differential diagnosis flatly inconsistent with epidemiological results has been deemed admissible, and citing *Joiner*'s "analytical gap" terminology, Judge McKay writes that "[w]e cannot allow the jury to speculate based on an expert's opinion which relies only on clinical experience in the absence of showing a consistent, statistically significant association between breast implants and systemic disease."[192] He dutifully lists the *Daubert* factors,[193] and suggests that the plaintiff's experts flunk at least two (peer review and publication, widespread acceptance);[194] but, as Professors Green and Sanders observe, this looks like window-dressing:[195] the core idea is that the plaintiff's expert testimony is inadmissible because, *given the epidemiological studies offered by the defendant, it is insufficient.*

But, again, as with relevance, other courts, and other judges, have taken a different line. I will focus on *Kuhn v. Wyeth* (2012),[196] where the plaintiffs alleged that their short-term use of Wyeth's hormone-replacement drug Prempro had caused their breast cancer, the defendants moved to have the plaintiffs' expert testimony excluded under FRE 702 and *Daubert*, and the magistrate judge to whom this evidentiary matter was referred granted their motion, leading to summary judgment for Wyeth.[197]

On appeal, however, Judge Wollman reverses this *Daubert* decision. He endorses Judge Becker's observation in *Paoli*: "[t]he standard for judging the evidentiary reliability of expert evidence is 'lower than the merits standard of correctness.'"[198] He continues: "[p]roponents of expert testimony need not demonstrate that the assessments of their experts are correct, and trial courts are not empowered 'to determine which of

191. Norris v. Baxter Healthcare Corp., 397 F.3d 878 (10th Cir. 2005).
192. *Id.* at 887.
193. *Id.* at 884.
194. *Id.* at 886.
195. Green & Sanders, *supra* note 172, at 226.
196. Kuhn v. Wyeth, Inc., 686 F.3d 618 (8th Cir. 2012).
197. *Id.*at 620–621.0
198. *Id.*at 625 (citing *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 744 (3rd Cir. 1994)).

several competing scientific theories has the best provenance.'"[199] And he cites the Advisory Committee's Notes on the 2000 revision of FRE 702: "When a trial court . . . rules that an expert's testimony is reliable, this does not necessarily mean that contradictory expert testimony is *un*reliable."[200] That, he writes (citing Justice Blackmun's observations in *Daubert*), is a matter to be sorted out by the finder of fact after cross-examination and presentation of contrary witnesses.[201]

Judge Wollman cites the "analytical gap" language of *Joiner*;[202] but tells us firmly that:

> Plaintiffs, as the proponents of Dr. Austin's testimony, however, do not necessarily have a burden to disprove the WHI [Women's Health Initiative] study's finding that short-term use of Prempro does not increase the risk of breast cancer. Instead, it is their burden to show that Dr. Austin arrived at his contrary opinion in a scientifically sound and methodological fashion.[203]

And he continues in the same vein with respect to the other studies Dr. Austin considered, again emphasizing the importance of the distinction between admissibility and weight.[204] True, toward the end of his ruling Judge Wollman writes that Dr. Austin's testimony is admissible "because the studies on which he relied were sufficient to support his opinion . . ";[205] but this obviously doesn't show that he is conducting a sufficiency analysis to determine whether Dr. Austin's testimony is admissible. On the contrary, he insists that while "[t]here may be several studies supporting Wyeth's contrary position," it is simply "*not the province of the court to choose between the competing theories when both are supported by reliable scientific evidence.*"[206] As the italicized phrase reveals, he assumes—as the Advisory Committee had said unmistakably clearly—that, in the sense explained in *Daubert*, there may be reliable evidence *both* for the plaintiff's theory, *and* for the defendant's competing explanation.

---

199. *Kuhn*, 686 F.3d at 625 (citing Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 15 (1st Cir. 2011) (quoting Ruiz-Troche v. Pepsi Cola of P.R. Bottling Co., 161 F.3d 77, 85 (1st Cir. 1998)).

200. *Kuhn*, 686 F.3d at 625 (emphasis added).

201. *Id.* at 625 (citing Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 596 (1993) ("Daubert I")).

202. *Id.*

203. *Id.*at 626.

204. *Id.*at 628–32.

205. *Id.*at 632.

206. Kuhn v. Wyeth, Inc., 686 F.3d 618, 633 (8th Cir. 2012) (emphasis added).

So there's yet another irony in *Daubert*. The Supreme Court granted certiorari in this landmark case to resolve disagreement across the federal courts about the status of *Frye* after the adoption of FRE 702; but, as we have seen, the conceptual fault line in *Daubert* has led to further disagreement, a new source of discord.

4. NORMATIVE PUZZLES

The main themes thus far have been descriptive and conceptual: that the contrast between the avowed intentions of the *Daubert* Court, and the consequences of its ruling—between its professed preference for admissibility and its more-restrictive results—is, in significant part, the consequence of the potential for slippage in those central concepts, relevance and reliability. But of course this leaves a tangle of normative questions unanswered. Are those more-stringent interpretations of relevance and reliability to be welcomed, or should they be resisted—and, in either case, on what grounds?

Opinions differ. A year after *Daubert*, summarizing some early decisions under the new evidentiary régime, Thomas J. Mack puzzled over how scientific testimony that was ruled admissible under *Daubert* could fail to be sufficient. Justice Blackmun's ruling, he suggested: ". . . carries the seemingly contradictory assumption that scientific testimony can be admissible as relevant and reliably grounded in scientifically valid reasoning and methodology and also be so "shaky" that it is insufficient to establish what it asserts."[207] This, he continued, could be read as an "inappropriate conflation" of admissibility and weight.[208]

Though he didn't articulate it specifically, Mack's argument already hinted at the problem that Prof. Imwinkelried would highlight a few years later, after *Daubert II*[209] and *Hall*:[210] that scientific testimony might be deemed *in*admissible *unless* it is sufficient: ". . . *Daubert II* and *Hall* are arguably authority that to be of enough assistance to be *admissible* under Federal Rule 702, standing alone purported scientific testimony must possess *sufficient* probative value to prove the fact in issue."[211] And, like Mack, Imwinkelried writes of the danger of a "conflation" of admissibility with sufficiency: "The reasoning of those courts about the second prong in the *Daubert* test is debatable and disturbing. In their

---

207. Thomas J. Mack, *Scientific Testimony after Daubert: Some Early Returns from Lower Courts*, 30 TRIAL 23, 30 (August 1994).

208. *Id.*

209. Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1314 (1995) ("Daubert II").

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

211. Imwinkelried, *supra* note 122, at 575.

exegesis of the helpfulness prong, the courts seemed to conflate admissibility and sufficiency analysis."[212]

Unlike Mack, however, who thought that *Daubert* "[might] not lead to much substantive change in the outcome of cases that turn on scientific testimony,"[213] Imwinkelried feared that enforcement of the principle that scientific testimony is admissible only if it is sufficient would "deal a tremendous blow to prosecutors," who could find themselves unable to introduce helpful but insufficient forensic testimony.[214] Perhaps needless to say, neither Mack's prediction nor Imwinkelried's has been borne out. *Daubert has* had an effect; but this effect has been felt much more in civil cases than in criminal,[215] notably in making it harder for plaintiffs in toxic tort cases to get their expert testimony admitted—perhaps because defendant companies in such cases have the resources to mount serious *Daubert* challenges, as parties now routinely do.[216]

More recently, Green and Sanders have taken a very different tack, defending some courts' elision of admissibility into sufficiency as a desirable development. Their main concern, they say, is simply to describe how the *Daubert* inquiry has evolved in toxic-tort litigation; but, they add, "normatively we applaud this reconceptualization of what *Daubert* is about . . . ." Why so? All they say explicitly is that this "will provide greater coherency and illumination to the process of deciding which cases should be submitted to a jury and which should not";[217] but their sub-title, "Controlling the Quality of Expert Witness Testimony in the United States," conveys the impression that they think this will encourage *better* expert testimony.

I wish the normative issues here were nearly as simple as they apparently seem, one way or the other, to these commentators. Sadly, they're not. Still, the suggestions and arguments we've been examining raise three kinds of issues worth exploring: first, about the interpretation of *Daubert*, etc.; second, about the underlying epistemological issues; and third, about matters of judicial economy, policy, and the like.

---

212. *Id.* at 583.

213. Mack, *supra* note 207, at 23.

214. Imwinkelried, *supra* note 122, at 580.

215. *See generally* Peter Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggestions for Reform*, 95 AM. J. PUB. HEALTH S107 (2005).

216. Green & Sanders, *supra* note 172, at 203.

217. Michael D. Green & Joseph Sanders, Admissibility versus Sufficiency: Controlling the Quality of Expert Witness Testimony in the United States (unpublished manuscript) (on file with author). This sentence has apparently been suppressed in the published (2016) version of their paper.

The first issue is raised by Mack's and Imwinkelried's talk of "conflation" or "confusion" of admissibility with sufficiency: is a more modest interpretation of relevance and reliability, or a more ambitious one, more faithful to FRE 702, to the *Daubert* trilogy, and to the underlying legal considerations? Here, I believe, the answer is clear: the modest understanding articulated by Judge Becker in *Paoli* and seconded by Judge Wollman in *Kuhn* is entirely consonant with the *Daubert* Court's observations about the "austerity" of the *Frye* Rule, the "preference for admissibility" articulated in the FRE, and the role of cross-examination and presentation of contrary evidence in exposing "shaky but admissible" expert testimony; and, of course, with the traditional distinction between admissibility and weight. More stringent interpretations, though they exploit a real elasticity in the key concepts in *Daubert*, look like—well, like a stretch. Moreover—even though, as we have seen, they include elements that may have encouraged more ambitious understandings—there is nothing in *Joiner* or *Kumho Tire* that *requires* the stronger interpretation; indeed, even in *Joiner* we read that "the Federal Rules of Evidence allow district courts to admit a somewhat broader range of scientific testimony than would have been admissible under *Frye*."[218] And, as we have also seen, the Advisory Committee's Notes on the 2000 revision of FRE 702 call unambiguously for a modest reading.[219]

But even if the more modest interpretation *is* more faithful to *Daubert*, etc., mightn't the more ambitious one have better consequences? Couldn't the unintended consequences of that fault line in *Daubert* be benign, even desirable? Well, I reply, certainly the unintended consequences of legal changes *can* be adaptive. But nothing follows about whether a more modest interpretation of relevance and reliability or, as Green and Sanders suggest, a more ambitious one, would result in less uncertainty and more predictability in decisions on the admissibility of expert testimony—much less about whether a weaker or a stronger interpretation would better control the quality of such testimony.

The claim about coherence seems to be ambiguous—and mistaken whichever way you take it. If the idea is that, if courts adopted more stringent interpretations, we could expect to get more consistent admissibility decisions in similar cases than we would if courts adopted more modest interpretations, the reply is surely that we could expect

218. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 142 (1997).

219. These Notes also make it abundantly clear that the reference in FRE to "sufficient" data is just a verbal variant on "enough" data or "adequate" data, and should *not* be taken as suggesting that reliability requires legal sufficiency.

more consistent results *either* way, whether courts more consistently took the more ambitious line, or more consistently took the more modest one. And if, on the other hand, the idea is that there is a kind of incoherence in supposing, as more modest interpretations must do, that *both* sides' expert testimony might be, in *Daubert*'s sense, reliable, the reply is that this ignores that fact that, when the science concerned is as yet unsettled, disagreements between equally-qualified, equally-serious, and equally-honest experts may be entirely reasonable given the limitations of the evidence available; it may be a matter of judgment what conclusion is better-warranted.[220]

What about the idea that more stringent interpretations would help ensure that parties produce better expert testimony? Discussing Judge Wollman's adoption of the modest interpretation of reliability in *Kuhn*, Green and Sanders suggest a quasi-epistemological argument: under *Daubert*, courts are to look to whether an expert's testimony was arrived at by the scientific method; and "Weight of Evidence Methodology," they continue—citing Prof. Cranor—is part of that method.[221] So, presumably, the idea is that Judge Wollman should have deemed the plaintiff's experts' testimony inadmissible because they hadn't taken account of *all* the evidence; which would have required them to give a satisfactory explanation of what was wrong with the defendants' contrary evidence.[222]

Unfortunately, Prof. Cranor's explanation of "weight of evidence methodology" was, to say the least, confusing;[223] and, in any case, I don't

---

220. This is not to say that this is subjective, simply a matter of taste; rather, it's to say that it is something that depends on the whole complex mesh of an expert's background beliefs, which naturally affects how much weight different experts give to this or that evidence.

221. Green & Sanders, *supra* note 172, at 232 (*citing* CARL CRANOR, TOXIC TORTS: SCIENCE, LAW, AND THE POSSIBILITY OF JUSTICE, 77–79, 136–39 (2006)).

222. *Id.*

223. Cranor writes of "inference to the best explanation," "diagnostic arguments," "diagnostic induction," "inductive arguments," and "differential diagnosis," apparently treating all these as equivalent. He tells us that such inferences may be stronger or weaker but not, like deductive inferences, valid or invalid. CARL CRANOR, TOXIC TORTS: SCIENCE, LAW, AND THE POSSIBILITY OF JUSTICE, 78 (2006). But he tells us very little about what makes such an inference more or less strong. According to Prof. Cranor's testimony in Milward, "weight of evidence methodology" (there apparently identified with "inference to the best explanation") has the following six steps: identify an association between exposure to substance S and the development of disorder D; consider a range of plausible explanations; rank them in order of plausibility; seek more evidence; consider all of the relevant available evidence; use your judgment to integrate it. Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 17–18 (1st Cir. 2011). But without substantial content given to the concepts of plausibility, relevance, and integration, this is empty advice.

believe there's any such "methodology."[224] Still, of course I agree that—in science as elsewhere—more comprehensive evidence is epistemologically better than less.[225] But *this* thought, though true, obviously doesn't answer the key questions presently at issue. Yes—assuming for the sake of argument that judges do an OK job of screening—making the standard for the admissibility of expert testimony more demanding should improve the quality of the expert testimony *presented to juries*. But this tells us nothing either way about whether, in the long run and on the whole, the legal system is likelier to get better expert testimony by having courts take a tougher approach to their screening task at the admissibility stage, so that fact-finders never hear what a judge deems inadequate, or by having judges take a more liberal approach to the task of screening expert testimony and letting the evidence be thrashed out at trial through cross-examination and the presentation of contrary witnesses. Perhaps needless to say, I don't know the answer to that question. Perhaps more importantly, I don't believe *anyone* does.

To be sure, the closer admissibility is pushed towards sufficiency, and so the more plaintiffs' expert testimony is found inadmissible, the more summary judgments we can expect; which will certainly save courts' time and energy. So, yes, more stringent interpretations would result in a gain in judicial economy. But not without cost: for one thing, the standard of review for decisions on the admissibility of expert testimony (abuse of discretion) is more deferential than the standard for review of summary judgment decisions (*de novo*); so this might mean a covert shift towards less rigorous review of summary judgment decisions.[226] Perhaps needless to say, I don't know whether, in the long run and on the whole—even if it does mean that some deserving plaintiffs will never have their cases heard, and perhaps that the potential dangers of some drugs or chemicals will never be explored as they ideally should be—this downside of tougher admissibility screening is more than outweighed by its benefits in streamlining a badly-overburdened tort system. I don't believe anyone knows the answer to *this* question, either.

---

224. *See* HAACK, *supra* note 120, at 226–27 & 235–37.

225. Indeed, according to the theory I developed in *Evidence and Inquiry*, comprehensiveness is one of the determinants of evidential quality. HAACK, EVIDENCE AND INQUIRY, *supra* note 79, at 117–39.

226. Presumably recognizing this, Green and Sanders suggest that, on appeal, decisions under the admissibility-as-sufficiency interpretation they favor be given a "hard look" under the abuse of discretion standard. Green & Sanders, *supra* note 172, at 237-38. This "concession" is, however—well, less than generous.

It's not that my sympathies lie, in principle, more with plaintiffs than with defendants. I don't doubt that plaintiffs' expert testimony in the kinds of toxic-tort case we have been exploring is sometimes pitifully weak—any more than I doubt that defendants' expert testimony is sometimes very selective and misleading. No: my main concern is that the tort system do a decent job both of its primary task, compensating deserving plaintiffs but not the undeserving, and of its secondary task, serving as backup to an inevitably fallible regulatory system by providing incentives to get dangerous products off the market without discouraging the production of useful and harmless ones. And that, I now see, is why I find the elision of reliability into sufficiency disturbing. If there's a defensible epistemological rationale for the adversarial system, it is that the thrashing-out of evidence at trial enables us, often enough, to reach factually sound verdicts.[227] But the elision of admissibility into sufficiency contributes to the growing tendency to pre-empt this process[228]—when no one knows whether the benefits are worth the cost.[229]

227. *See* Susan Haack, *Epistemology Legalized: Or, Truth, Justice, and the American Way*, 49 AM. J. JURIS. 43 (2004), *reprinted in* EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN THE LAW 27 (2014); Susan Haack, Justice, Truth, and Proof: Not So Simple, After All, presentation at the Cátedra de Cultura Jurídica at the Universitat de Girona, Spain (Jan. 1, 2015), *forthcoming in* Spanish translation in DEBATIENDO CON MICHELE TARUFFO (Jordi Ferrer Beltrá & Carmen Vázquez, eds.) (English version on file with Author).

228. *See* Haack, Justice, Truth, and Proof, *supra* note 227 (English manuscript pages 20–24).

229. My thanks to Mark Migotti, for helpful comments on more than one draft; to Erica Beecher-Monas and Stephen Calkins, for the helpful questions that prompted notes 96 and 76 respectively; to Joseph Sanders, for helpful correspondence; to Pamela Lucken and A.J. Blechner, for help in locating relevant (and reliable!) material; to Vitor de Paula Ramos, for teaching me that marvelous Portuguese idiom for "I have a problem"; and, of course, to the Cohen family for their generosity in endowing this lecture series.