

2008

# Proving Causation: The Holism of Warrant and the Atomism of *Daubert*

Susan Haack

*University of Miami School of Law*, [shaack@law.miami.edu](mailto:shaack@law.miami.edu)

Follow this and additional works at: [https://repository.law.miami.edu/fac\\_articles](https://repository.law.miami.edu/fac_articles)



Part of the [Evidence Commons](#)

---

## Recommended Citation

Susan Haack, *Proving Causation: The Holism of Warrant and the Atomism of Daubert*, 4 *J. Health & Biomed. L.* 253 (2008).

This Article is brought to you for free and open access by the Faculty and Deans at University of Miami School of Law Institutional Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Miami School of Law Institutional Repository. For more information, please contact [library@law.miami.edu](mailto:library@law.miami.edu).

## Proving Causation: The Holism of Warrant and the Atomism of *Daubert*<sup>1</sup>

Susan Haack<sup>2</sup>

*The Consilience of Inductions* takes place when an Induction, obtained from one class of facts, coincides with an Induction, obtained from a different class. This Consilience is a test of the truth of the Theory in which it occurs. — William Whewell<sup>3</sup>

As my title indicates, this article focuses on causation evidence in toxic-tort litigation; and as my sub-title suggests, it makes two main arguments, the first epistemological and the second legal. The epistemological argument is that, under certain conditions, a congeries of evidence warrants a conclusion to a higher degree than any of its components alone would do; the legal argument, interlocking with this, is that our evidence law imposes a kind of atomism that can actually impede the process of arriving at the conclusion most warranted by the evidence – the effects of which have been especially salient to causation evidence in toxic-tort cases.

Section 1 will set the stage by looking at some cases where the epistemological issue to be tackled here came explicitly to courts' attention; section 2 will develop the epistemological argument, first in a general form, and then as it applies to the kinds of causation evidence typically encountered in toxic-tort litigation; section 3 will rely on this account to answer some of the epistemological questions about causation evidence that have been at issue in such cases; and section 4 will develop the legal argument, showing that, ironically enough, *Daubert's* requirement that expert testimony be reliable may

---

<sup>1</sup> © Susan Haack 2008.

<sup>2</sup> Distinguished Professor in the Humanities, Cooper Senior Scholar in Arts and Sciences, Professor of Philosophy, Professor of Law, University of Miami.

<sup>3</sup> William Whewell, *Philosophy of the Inductive Sciences*, in *Selected Writings of William Whewell*, 212-259, 257 (Yehuda Elkana, ed., 1984) (The word “consilience,” which I believe Whewell introduced, derives from the Latin “con” and “silere,” “jumping together”).

sometimes stand in the way of an adequate assessment of the reliability of causation evidence.

### 1. Setting the Stage

Mary Virginia Oxendine was born in 1971. Her right forearm was foreshortened, and she had only three fingers, fused together, on her right hand. Believing that their daughter's birth defects had been caused by her mother's taking Bendectin for morning-sickness while pregnant with Mary, the Oxendines sued the manufacturers, Merrell Dow Pharmaceuticals.<sup>4</sup>

At the first jury trial, Dr. Alan Done testified for the Oxendines that certain anti-histamines are known to have teratogenic effects on animals, and that one ingredient of Bendectin is the anti-histamine doxylamine succinate; that animal studies conducted by Merrell Dow found small limb alterations in the fetuses of pregnant rabbits given Bendectin – alterations the company scientists disregarded as insignificant – as well as miscarriages which, Dr. Done believed, may have occurred because the babies were malformed; that *in vitro* studies conducted by the National Institutes of Health found that Bendectin interfered with the development of limb-bud cells; and that the data from an epidemiological study conducted for Merrell Dow by Drs. Bunde and Bowles, when adjusted to exclude Canadian subjects (who could have bought the drug without a prescription), revealed that mothers who took Bendectin had a 30% greater risk of having a deformed baby.<sup>5</sup> Dr. Done explained that his belief that Mary Oxendine's birth defects had been caused by the Bendectin her mother had taken during the period of pregnancy in which fetal limbs were forming was based, not on any one of these studies or any one of these lines of evidence by itself, but on *all* of the various pieces of evidence to which he testified, *taken together*.<sup>6</sup>

---

<sup>4</sup> *Oxendine v. Merrell Dow Pharm., Inc.* 506 A.2d 1100 (D.C. 1986), *on remand*, 563 A.2d 330 (D.C. 1989), *cert. denied*, 493 U.S. 1074 (1990), *on remand*, 593 A.2d 1023 (D.C. 1991), *on remand*, 649 A.2d 825 (D.C. 1994), *on remand*, Civ. No. 82-1245, 1996 WL 680992 (D.C. Super. Oct. 24, 1996). The description of Ms. Oxendine's birth defects comes from *Oxendine*, 506 A.2d at 1103.

<sup>5</sup> *Oxendine*, 506 A.2d at 1104-109 (reporting part of Dr. Done's testimony). What I have given in the text is obviously only a very sketchy summary of Dr. Done's testimony; he was on the witness stand for three and a half days, and the transcript of his testimony runs to almost 600 pages. *Id.* at 1108.

<sup>6</sup> *Oxendine*, 506 A.2d at 1108 (reiterating that "[Dr. Done] conceded his inability to conclude that Bendectin is a teratogen on the basis of any of the individual studies which he discussed, but he also made clear that all of these studies must be viewed together, and that, so viewed, they supported his conclusion").

In 1983, at the first trial, a jury awarded the Oxendines \$750,000 in compensatory damages. But, overriding this decision, writing that “it is clear . . . that *no conclusion one way or the other can be drawn from any of the above relied upon bases* respecting whether Bendectin is a human teratogen,”<sup>7</sup> the court granted the defendant’s motion for summary judgment notwithstanding the verdict. The Oxendines appealed; and the Court of Appeals reversed and remanded with instructions to reinstate the jury verdict, finding that the trial court had erred in emphasizing Dr. Done’s acknowledgment that none of the individual studies to which he testified was sufficient by itself to establish causation, and “failing to consider [his] testimony that all of the studies, taken in combination, did support such a finding.” Associate Judge Terry continued:

Like the pieces of a mosaic, the individual studies showed little or nothing when viewed separately from one another, but *they combined to produce a whole that was greater than the sum of its parts*: a foundation for Dr. Done’s opinion that Bendectin caused appellant’s birth defects.<sup>8</sup>

Of course, this was not the end of the *Oxendine* story: in fact the case went to the D.C. Court of Appeals three more times before it was finally resolved in 1996. On remand, Merrell Dow moved for a new trial, claiming that Dr. Done had misrepresented his credentials;<sup>9</sup> and in 1988 this motion was granted. The Oxendines appealed again and in 1989, finding that the trial judge had erred in granting a new trial, the Court of Appeals reversed again, once more ordering the trial court to reinstate the original verdict.<sup>10</sup> Back at the trial court, the Oxendines asked the court to enter a judgment affirming the verdict, but Merrell Dow appealed once more; and in 1991 the Court of Appeals ruled that the trial court could not enter a final, unappealable judgment on compensatory damages until the punitive-damages stage of the trial was completed.<sup>11</sup> In

---

<sup>7</sup> *Oxendine*, 506 A.2d at 1103 (emphasis added).

<sup>8</sup> *Oxendine*, 506 A.2d at 1110 (emphasis added) (determining that the trial court’s summary judgment was in error, because when the evidence was viewed as a whole, it was not appropriate to conclude that no reasonable juror would find for the appellant).

<sup>9</sup> *Oxendine*, 563 A.2d at 332 (reporting that on May 3 and May 11, 1983, Dr. Done had testified that he was a member of the Wayne State Medical School Faculty, when in fact he had submitted a letter of resignation on April 24, which was accepted by the Dean on April 29th; and listing four other respects, in addition to his position at Wayne State Medical School, in which Dr. Done falsely represented his credentials at trial).

<sup>10</sup> *Oxendine*, 563 A.2d at 331 (finding that the motions judge did not abuse his discretion in finding that the motion to vacate was timely, but that he did err in vacating the judgment and granting a new trial). *Id.* at 338 (reversing and ordering the trial court to reinstate the jury verdict).

<sup>11</sup> *Oxendine*, 593 A.2d at 1023 (reversing award for compensatory damages before punitive damage stage of trial was completed).

1993, the Oxendines withdrew their claim for punitive damages, and moved for the verdict on compensatory damages to be affirmed; and Merrell Dow asked the trial court to reconsider the original verdict, this time on the grounds that new studies published since the first trial exonerated Bendectin. The trial court, declining to consider these new studies, entered a judgment reaffirming the original jury verdict. Merrell Dow appealed again; and in 1994, acknowledging that “reopen[ing] the trial’s determination of scientific truth” was at odds with the legal concern for finality,<sup>12</sup> and therefore setting a high standard for Merrell to prevail, the Court of Appeals remanded yet again – as the court says, reluctantly, and evidently expecting that the case would be quickly resolved in favor of the Oxendines.<sup>13</sup>

But in 1996– now taking into account the new studies Merrell Dow presented,<sup>14</sup> the decisions in numerous other Bendectin cases concluded since the original trial,<sup>15</sup> and

---

<sup>12</sup> *Oxendine*, 649 A.2d at 831-32 (stressing importance of finality in the legal system). See also Susan Haack, *Irreconcilable Differences? The Troubled Marriage of Science and Law*, L. & CONTEMP. PROBS. (forthcoming 2008) (arguing in part that there is tension between the open-ended fallibilism of scientific inquiry and the legal desideratum of finality).

<sup>13</sup> *Oxendine*, 649 A.2d at 827 (finding that “we are reluctantly compelled to remand for further limited consideration”); see also *id.* at 834 (Associate Judge Schwelb, concurring, commenting that “[t]he delays to date . . . have *already* done intolerable damage [T]his is not 1982 or 1984 or even 1990 . . . Given where we are today, considerations of finality have become so compelling that . . . nothing short of an extraordinarily persuasive proffer by Merrell Dow would warrant . . . further delaying Ms. Oxendine’s recovery.”).

<sup>14</sup> *Oxendine*, 1996 WL 680992 at 14-21 (reporting that Merrell had presented 2 post-1983 meta-analyses of epidemiological data on Bendectin (Einarson et al., 1988; McKeigue et al., 1994), and 14 epidemiological studies (Golding, 1983; Mitchell, 1983; Aselton-Jick, 1984; Hearey, 1984; McCredie, 1984; Winship, 1984; Elbourne, 1985; Aselton-Jick, 1985; Zieler, 1985; Jedd, 1988; Shiono, 1989; Erickson, 1991; McDonald, 1991; Khoury, 1994)). Plaintiffs’ counsel argued that these studies, where they were relevant, were flawed; e.g., that the 1991 Erickson study omitted crucial safeguards such as “critical times” (presumably, the period of pregnancy in which subjects took Bendectin), but the court downplayed these criticisms as “counsel’s critique of a scientific study, rather than a contrary scientific study or expert evaluation.” *Id.* at 15 (citing and dismissing counsel’s arguments).

<sup>15</sup> *Oxendine*, 1996 WL 680992 at 4-7 (listing eight federal cases concluded in favor of Merrell: *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311 (9th Cir. 1995), *cert. denied*, 516 U.S. 869 (1995); *Turpin v. Merrell Dow Pharm., Inc.*, 959 F.2d 1349 (6th Cir. 1992), *cert. denied*, 506 U.S. 826 (1992); *Wilson v. Merrell Dow Pharm., Inc.*, 893 F.2d 1159 (10th Cir. 1990); *Ealy v. Richardson-Merrell, Inc.*, 897 F.2d 1159 (D.C. Cir. 1990), *cert. denied*, 498 U.S. 950 (1990); *De Luca v. Merrell Dow Pharm., Inc.*, 911 F.2d 941 (3rd Cir. 1990); *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988), *cert. denied*, 493 U.S. 882 (1989); *Brock v. Merrell Dow Pharm., Inc.*, 874 F.2d 307 (5th Cir. 1989); *Lynch v. Merrell-National Laboratories*, 830 F.2d 1190 (1st Cir. 1987). The court also mentions *Blum* and *Hawner*, but observes that both are on appeal. *Oxendine*, 1996 WL680992 at 7. Both were eventually resolved in favor of the

actions of the FDA<sup>16</sup> and the Canadian government<sup>17</sup> – the trial court found that this high standard was met, and found in favor of Merrell Dow.<sup>18</sup>

The best known of the other Bendectin cases cited was, of course, *Daubert*, on which the U.S. Supreme Court had given its landmark ruling in 1993;<sup>19</sup> and which had been finally resolved a year before *Oxendine*, when Judge Kozinski affirmed the trial court's summary judgment for Merrell Dow.<sup>20</sup> Jason Daubert's birth defects were similar to Mary Oxendine's,<sup>21</sup> and his parents, like hers, believed these defects had been caused by Bendectin; but legally *Daubert* followed a different path from *Oxendine*. In 1989 the District Court had granted Merrell Dow's motion for summary judgment after excluding the Dauberts' proffered expert witnesses on the grounds that scientific evidence is

---

defendants. See *Blum ex rel. Blum v. Merrell Dow Pharm., Inc.*, 764 A.2d 1, 4-8 (Pa. 2000) (affirming Superior Court's decision in favor of Merrell); *Merrell Dow v. Havner*, 953 S.W.2d 706, 708 (Tex. 1997) (reversing the court of appeals and finding in favor of Merrell). The two names of the defendant company – Richardson-Merrell, Merrell-Dow – reflect changes in ownership over the relevant period. See JOSEPH SANDERS, *BENDECTIN ON TRIAL: A STUDY OF MASS TORT LITIGATION* 213-14 (1998) (describing the history of the company); see also MICHAEL GREEN, *BENDECTIN AND BIRTH DEFECTS* (1996) (describing the history of Bendectin litigation).  
<sup>16</sup> *Oxendine*, 1996 WL 680992 at 23 (referring to a monograph on over-the-counter anti-histamine drugs issued by the FDA in 1994 that examined doxylamine succinate and concluded that it was safe to include as an ingredient of such anti-histamines).

<sup>17</sup> *Oxendine*, 1996 WL 680992 at 23 (reporting that in 1988, the consultants for the Special Advisory Committee on Reproductive Physiology to the Health Protection Branch of the Canadian government concluded that Bendectin should not be withdrawn from the Canadian market and that the warning label should be modified in light of the lack of evidence of an association with birth defects. *But see id.* at 23 n.45 (noting that plaintiffs' counsel pointed out that the members of the Canadian panel "were tied to Merrell – a fact of which the Canadian government was not aware.")).

<sup>18</sup> *Id.* at 34. In telling the tangled tale of this long-running legal saga I have relied in part on the history recounted in Joseph Sanders, *Science, Law and the Expert Witness*, L. & CONTEMP. PROBS. (forthcoming 2008). Earlier, Prof. Sanders had speculated, very plausibly, that Merrell Dow expended so much time and money on its defense in *Oxendine* "in order to maintain an unblemished record in the Bendectin litigation. Even one final plaintiff verdict might make it more difficult to argue for a summary judgment in other cases." See SANDERS, *supra* note 15, at 30.

<sup>19</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) (determining standard for admitting expert scientific testimony at a federal trial).

<sup>20</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1322 (9th Cir. 1995) (affirming summary judgment).

<sup>21</sup> See Natalie Angier, *High Court to Consider Rules on Use of Scientific Evidence*, N. Y. TIMES, Jan. 2, 1993, at 1, available at ProQuest Historical Newspapers The New York Times (1851-2003) (reporting that "Jason Daubert, of San Diego, was born 19 years ago with only two fingers on his right hand and without a lower bone on his right arm.").

admissible only if it is “sufficiently established to be generally accepted in the field to which it belongs,”<sup>22</sup> and finding that, since none of the numerous published epidemiological studies had found a statistically significant association between Bendectin and birth defects, the Dauberts’ experts’ opinions were *not* generally accepted in the field to which they belonged, and hence not admissible. The Court of Appeals for the 9<sup>th</sup> Circuit affirmed, specifically citing *Frye*.<sup>23</sup> And because of this reliance on *Frye* – almost unprecedented in a civil case<sup>24</sup> – the Supreme Court granted certiorari, to determine whether or not *Frye* had been superseded when the Federal Rules of Evidence (FRE) were adopted in 1975.

An amicus brief from Kenneth Rothman and other epidemiologists raised several important epistemological issues; the lower courts’ analyses in *Daubert*, these amici argued, put too much weight on whether studies were statistically significant, over-estimated the importance of peer-reviewed publication<sup>25</sup> – and, most to the present purpose, they “foreclose[d] the use of valid inferences that may be drawn from the combination of

---

<sup>22</sup> See *Daubert v. Merrell Dow Pharm., Inc.*, 721 F. Supp. 570, 572 (S. D. Cal. 1989) (citing *United States v. Kilgus*, 571 F.2d 508, 510 (9th Cir. 1988)) (citing *United States v. Brown*, 557 F.2d 541, 556 (6th Cir. 1977)) (citing *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923)).

<sup>23</sup> See *Daubert v. Merrell Dow Pharm., Inc.*, 951 F.2d 1128, 1129-1131 (9th Cir. 1991) (citing *Frye*, 293 F. at 1014).

<sup>24</sup> Kenneth J. Cheseboro, *Galileo’s Retort: Peter Huber’s Junk Scholarship*, 42 AM. U. L. REV. 1637, 1695 (1993) (reporting that “there was not a single case decided by the federal appellate courts prior to 1975 that applied the *Frye* rule in a civil case of any kind. As of April 7, 1993, only three such decisions had been reported, two of which were decided in 1991”). The three decisions were: *Christopherson v. Allied-Signal Corp.*, 939 F.2d 1106, 1115-16 (5th Cir. 1991), *cert. denied* 503 U.S. 912 (1992); *Daubert*, 951 F.2d 1128 (9th Cir. 1991); *Barrel of Fun, Inc. v. State Farm Fire and Casualty Co., Inc.*, 739 F.2d 1028, 1031 (5th Cir. 1988). See Cheseboro, *supra*, at 1695 n.264. Whether *Christopherson* really “relies” on *Frye* might be questioned, since the court lists four considerations, of which *Frye* is only one. See *Christopherson*, 939 F.2d at 1110. When the Supreme Court denied cert. in 1992, however, Justice White, with Justice Blackmun, dissented, arguing that the question, whether *Frye* had been superseded by the FRE, “is an important and recurring issue.” *Christopherson v. Allied Signal Corp.*, 503 U.S. 912, 913 (1992) (White, J., dissenting) (contending cert. should be granted). *Barrel of Fun*, which more unambiguously relies on *Frye*, was a fire-insurance fraud case in which the excluded evidence involved a “psychological stress evaluation” of proffered testimony, which the court held to be essentially similar to polygraph evidence, which was the kind of evidence at issue in *Frye*. See *Barrel of Fun, Inc.*, 739 F.2d at 1029 (seeing the evidence at issue as essentially similar to the excluded evidence in *Frye*).

<sup>25</sup> See also Brief for Petitioners, Daryl E. Chubin, Ph.D et al. as Amici Curiae, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) (dealing primarily with peer review); see also Susan Haack, *Peer Review and Publication: Lessons for Lawyers*, 36 STETSON L. REV. 789 (2007) (distinguishing broad and narrow concepts of peer review and exploring their role in *Daubert* and subsequent litigation).

*many studies, even when none of the studies standing alone would justify such inferences.*<sup>26</sup> But the Supreme Court's ruling – that *Frye* had been superseded, but that FRE 702 still required that courts screen proffered expert testimony for reliability as well as relevance – does not pick up this theme. However, Justice Blackmun's ruling had continued by stressing that in screening for reliability courts should look, not to an expert's conclusions, but to his "methodology."<sup>27</sup> And so, when *General Electric Co. v. Joiner*<sup>28</sup> came to the Supreme Court in 1997, the dispute over the question of the joint weight of combined causation evidence was couched in terms of the parties' rival experts' "methodologies."

Robert Joiner, who had worked for many years as an electrician for a municipality in Georgia, was diagnosed with small-cell lung cancer in 1991; he was only 37. Believing that his cancer had been promoted by his exposure to polychlorinated biphenyls (PCBs) contaminating the insulating oil in the transformers his job required him to disassemble and repair, he sued the manufacturer, General Electric Co. (G.E.). Mr. Joiner's attorneys had proffered a number of expert witnesses, who proposed to testify to a variety of toxicological, *in vitro*, *in vivo*, and epidemiological studies; arguing that, *taken together*, this congeries of evidence would be sufficient to establish causation. These experts, they explained, used "weight of evidence methodology" – the same methodology the Environmental Protection Agency used in assessing carcinogenic risk, and the same methodology G.E.'s own experts used in this very case. G.E., however, denying this imputation, argued that what Joiner's attorneys presented as reputable scientific methodology was actually nothing more than the "faggot fallacy": the mistake of supposing that a pile of weak evidence, if it is large enough, is magically transmuted into strong evidence.<sup>29</sup>

The District Court, focusing one-by-one on (some of) the individual studies to which Joiner's experts appealed, excluded Joiner's expert testimony as inadmissible, and granted G.E.'s motion for summary judgment. But the Court of Appeals reversed, holding that where, as in this case, exclusion of expert testimony is outcome-determinative, appellate review should be especially stringent; and, moreover, found Joiner's experts' methodology scientifically acceptable:

---

<sup>26</sup> Brief for Petitioners, Professor Kenneth Rothman et al., as Amici Curiae, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 570 at \*10 (1993) (No. 92-102) (emphasis added).

<sup>27</sup> *Daubert*, 509 U.S. at 592-3 (applying Rule 702 requires "a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether the reasoning or methodology properly can be applied to the facts at issue.").

<sup>28</sup> *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997).

<sup>29</sup> See Brief for Petitioners at \*49, *Gen. Elec. v. Joiner*, 522 U.S. 136 (1997).



Opinions of any kind are derived from individual pieces of evidence, each of which by itself might not be conclusive, but *when viewed in their entirety are the building blocks of a perfectly reasonable conclusion*, one reliable enough to be submitted to a jury. . . .<sup>30</sup>

The Supreme Court granted certiorari, to determine the standard of review for such evidentiary rulings. Ruling unanimously that the proper standard of review remained abuse of discretion, the *Joiner* Court sidestepped Joiner's argument about "weight of evidence methodology" with the brisk observation that methodology and conclusions "are not entirely distinct from one another," and that a court may reasonably conclude that there is "simply too great an analytical gap between the data and the opinion proffered." And then, briefly reviewing (some of) the testimony that Joiner's experts would have given had they been admitted, the court found, almost unanimously, that the District Court had not abused its discretion in excluding Joiner's experts.<sup>31</sup>

But on this last point there was one dissenter, Justice Stevens, who argued that it would have been better to have remanded the case to the Appeals Court for reconsideration under the appropriate standard of review. Joiner's experts had referred to numerous studies, he points out, only one of which is in the record, and only six of which were ever considered by the District Court; moreover, he continues, the majority view on the question of reliability, which required it to play down the distinction of methodology and conclusions, is "arguably not faithful to . . . *Daubert*."<sup>32</sup> (Indeed: after all, the distinction of methodology vs. conclusions, which the majority rather casually sets aside in *Joiner*, was front-and-center in *Daubert*).<sup>33</sup> And, like the Court of Appeals, Justice Stevens believes there is merit in Joiner's experts' epistemological argument:

It is not intrinsically "unscientific" for experienced professionals *to arrive at a conclusion by weighing all available scientific evidence* – this is not the

---

<sup>30</sup> *Joiner v. G.E.*, 78 F.3d 524, 532 (11th Cir. 1996) (emphasis added). G.E.'s attorneys claim that "[a]lmost [these] very words have been cited by scientists and scholars as violating the methodology of science." See Brief for Petitioners at \*49, Gen. Elec. V. *Joiner*, 522 U.S. 136 (1997) (citing PETR SKRABANEK & JAMES MCCORMICK, *FOLLIES AND FALLACIES IN MEDICINE* 35 (Prometheus Books 1990) (characterizing the "faggot fallacy").

<sup>31</sup> See *Joiner*, 522 U.S. at 146-147 (holding that abuse of discretion is the applicable standard, and that the district court did not abuse its discretion in excluding Joiner's experts).

<sup>32</sup> *Id.* at 152 (Stevens, J., dissenting in part).

<sup>33</sup> *Daubert*, 509 U.S. at 595 ("The focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.").

sort of “junk science” with which Daubert was concerned. After all, as Joiner points out, the Environmental Protection Agency (EPA) uses the same methodology to assess risks, albeit using a somewhat different threshold . . .<sup>34</sup>

Of course, whether, and if so how, a compilation of pieces of evidence none of which is sufficient by itself to warrant a causal conclusion to the required degree of proof might do so jointly is a question that arises over and over in toxic-tort cases, though not usually as explicitly as in *Oxendine* and *Joiner*.<sup>35</sup> The epistemological puzzle comes out particularly vividly in the first case described here, in Dr. Done’s testimony in *Oxendine*: the structure-activity toxicological evidence is not sufficient to make the case for causation, he acknowledges, nor is the evidence from *in vitro* studies, nor the evidence from animal studies, nor his statistical re-analyses. But put them all together, however, he continues, and somehow – presto! – they amount to proof of causation.<sup>36</sup> But *how*, exactly? He doesn’t say; and neither, but for his nice metaphor of a mosaic, does Judge Terry. And so far as I know, this issue has yet to be satisfactorily resolved. The purpose of the next section is to fill this “analytical gap.”<sup>37</sup>

## 2. The Epistemological Argument

The first thing to notice is that, while up to now we have been approaching it from the perspective of causation evidence in toxic-tort cases, where legally it has been especially salient, this epistemological question is really quite general, arising in virtually every area of inquiry.<sup>38</sup>

---

<sup>34</sup> *Joiner*, 522 U.S. at 153 (emphasis added).

<sup>35</sup> *But see e.g.*, *Castillo v. E.I. Dupont de Nemours*, 854 So. 2d 1264, 1272 (Fla. 2003) (reporting that “[Dr. Van Velzen] repeatedly asserted that he used the in-vitro testing as one source of data, in conjunction with other reliable data, to reach the conclusion. He testified that *the consideration of all the data together is a commonly accepted scientific practice.*”) (emphasis added). I note for the record that Florida is a *Frye* state, and that the standard of review for *Frye* rulings is *de novo*.

<sup>36</sup> *See Oxendine v. Merrell Dow Pharm., Inc.*, 506 A.2d 1100, 1108 (D.C. 1986) (reporting that “[t]hroughout his testimony, [Dr. Done] repeatedly stated that his opinion was based not on any single study or type of evidence, but on four different types of scientific data viewed in combination”).

<sup>37</sup> These examples will recur throughout the paper; so perhaps it is necessary for me to say right away that my argument is *not* that Bendectin causes birth defects, or that PCBs cause small-cell lung cancer – or, of course, that they do not. Even if I had all the evidence – which, obviously, I do not – I would not be competent to make such judgments.

<sup>38</sup> *See Rothman, supra* note 26, at 10 (noting that “[t]his commonsense observation is not novel or controversial.”).

Think, for example, of that meteorite discovered in Antarctica in 1984 and believed, on the basis of the gases it gives off when heated, to have come from Mars about 4 billion years ago. A chemist at Stanford discovered that the meteorite contained molecules of polycyclic aromatic hydrocarbons (PAHs), which are found not only in diesel exhaust and soot, but also in decomposed organic matter; and other scientists discovered that the crystals of carbonate in the meteorite were shaped like cubes and teardrops, like those formed by bacteria on earth. By 1997, Dr. David Mackay of the Johnson Space Center was ready to say, in an interview for Newsweek, that “[w]e have these lines of evidence. None of them in itself is definitive, but taken together the simplest explanation is early Martian life,”<sup>39</sup> and as more evidence came in over a decade of so of further research, this conclusion has become more firmly warranted.<sup>40</sup> Nor is this an isolated example; on the contrary, with respect to virtually any well-warranted scientific claim of any importance – that DNA is the genetic material,<sup>41</sup> for example, or that species evolve through a process of natural selection<sup>42</sup> – the evidence is a complex mesh of interwoven elements.

Nor, for that matter, is this reliance on many intersecting lines of evidence confined to the sciences. Think, for example, of a historian relying on archeological and on documentary evidence (and perhaps also on the scientific theory underlying

---

<sup>39</sup> See Sharon Begley & Adam Rogers, *War of the Worlds: There Are No Little Green Men on Mars. But There Are Some Very Hostile Fellows on Earth Debating Whether There was Life on the Red Planet*, NEWSWEEK, Feb. 10, 1997, at 56-58 (emphasis added).

<sup>40</sup> See generally Thomas H. Maugh III, *Probe Enters Mars Orbit*, L.A. TIMES, Mar. 11, 2006, at A12 (reporting that it is now known that there was once water on Mars, and that a second Martian meteorite also contains what may be Martian fossils). See also Michael Hanlon, *Is This Proof of Life on Mars? The Meteorite That May Finally Have Resolved the Great Mystery*, DAILY MAIL, Feb. 10, 2006, at 40.

<sup>41</sup> In 1944, when Oswald Avery published the report of the experiments that are now recognized as having established that DNA, not protein, is the genetic material, he was unwilling to draw the conclusion in print, and it was not generally accepted until after Hershey and Chase's experiments, published in 1952. See Oswald T. Avery et al., *Studies of the Chemical Nature of the Substance Inducing Transformation in Pneumococcal Types*, 79 J. OF EXPERIMENTAL MEDICINE 137 (1944); A. D. Hershey & Martha Chase, *Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage*, 36 J. OF GENERAL PHYSIOLOGY 39 (1952). By 1953, when Watson and Crick published their paper on the structure of DNA, the role of DNA in heredity was only very imperfectly understood, and according to Crick only in the 1980s was the conclusion firmly established. See James D. Watson and Francis Crick, *Molecular Structure of Nucleic Acids: A Structure for Deoxyribonucleic Acid*, 171 NATURE 737 (1953); see also FRANCIS CRICK, *WHAT MAD PURSUIT: A PERSONAL VIEW OF SCIENTIFIC DISCOVERY* 7 (1988).

<sup>42</sup> See *Understanding Evolution: Your One-Stop Source for Information about Evolution*, available at <http://evolution.berkeley.edu> (providing a helpful summary of this extraordinarily extensive and varied evidence when you click on the link headed: *What Is the Evidence for Evolution?*).

techniques for dating remains, or for identifying the paper on which or the ink with which a document is written), or on a combination of written records and the testimony of still-living witnesses. In fact, this kind of reliance on a whole mesh of evidence is ubiquitous – the rule, not the exception. It is commonplace in everyday life: when, for example, after reading a startling story in a newspaper, I buy a different paper, or turn on the television news, to check whether other sources confirm it.<sup>43</sup> And this reliance on a combination of lines of evidence is familiar in many legal contexts too: when, for example, we ask a jury to arrive at a conclusion based on the testimony of eye-witnesses and of a psychologist testifying to the circumstances in which eye-witnesses are more, or less, reliable, or on forensic evidence and testimony about the error-rate of this laboratory, and so on.

Because the epistemological question at issue is quite general, we need a general answer. And warrant is clearly a matter of degree (as I took for granted in describing the hypothesis that there was early Martian bacterial life as weakly warranted a decade ago, and significantly more strongly warranted by now), so the answer needs to explain, first, what factors determine whether, and if so to what degree, evidence warrants a conclusion; and, second, under what conditions those factors work in such a way as to enhance degree of warrant when diverse pieces of evidence are combined. My answer will call on the account of the structure of evidence and the determinants of degree of warrant that I presented in *EVIDENCE AND INQUIRY*<sup>44</sup> and amplified and refined in *DEFENDING SCIENCE*.<sup>45</sup>

Evidence ramifies, rather as entries in a crossword puzzle do; and my account is informed by this analogy. How reasonable a crossword entry is depends on how well it fits with the clue and any already-completed intersecting entries; how reasonable those entries are, independent of the one in question; and how much of the crossword has been completed. Similarly, I suggest, how warranted a conclusion is (or, as we might put it more idiomatically, how likely the evidence makes it that the conclusion is true) depends on three factors:

---

<sup>43</sup> On a recent visit to Spain, intrigued by their names, I bought copies of both of the two newspapers published in Murcia: *LA VERDAD* (“TRUTH”) and *LA OPINIÓN*. (Friends told me that *LA VERDAD* was a very conservative publication, *LA OPINIÓN* more liberal.) Both carried the same front-page story, of a woman strangled in the center of the town.

<sup>44</sup> SUSAN HAACK, *EVIDENCE AND INQUIRY: TOWARDS RECONSTRUCTION IN EPISTEMOLOGY* 73-94 (1993) (2nd ed. forthcoming 2009).

<sup>45</sup> SUSAN HAACK, *DEFENDING SCIENCE –WITHIN REASON: BETWEEN SCIENTISM AND CYNICISM* 57-91 (2003).

- (i) how strong the connection is between the evidence and the conclusion – *supportiveness*;
- (ii) how solid the evidence itself is,<sup>46</sup> independent of the conclusion – *independent security*; and
- (iii) how much of the relevant evidence the evidence includes – *comprehensiveness*.

I note that, though we often speak of degree of supportiveness in terms of how likely this or that evidence makes this conclusion, or of degree of warrant in terms of how likely it is that the conclusion is true, these are *epistemic* likelihoods; they cannot properly be construed as statistical probabilities. Indeed, given the multidimensional character of the determinants of evidential quality, there is no guarantee even of a linear ordering of degrees of warrant, let alone a realistic possibility of assigning (meaningful) numbers to them.<sup>47</sup>

I also note that these three factors are not quite symmetrical. Supportiveness is directly correlated with degree of warrant; i.e., the more supportive the evidence is of a conclusion, the better warranted the conclusion (as a crossword entry is more reasonable the better it fits with the clue and other completed entries). But the connection between independent security and warrant is a bit more complicated. The more independently secure evidence *for* a conclusion is, the more warranted the conclusion; but the more independently secure the evidence *against* a conclusion is, the less warranted the conclusion (as, in a crossword, the fact that our answer to 4 down fits with our answer to 2 across is the more reassuring the more confident we are that 2 across is right; but if our answer to 4 down *doesn't* fit with 2 across, this is less troubling the *less* confident we are that 2 across is right). Similarly, the more comprehensive evidence *for* a conclusion is, the better warranted the conclusion; but if making the evidence more comprehensive also makes it less positive, the increase in comprehensiveness lowers the degree of

---

<sup>46</sup> See HAACK, DEFENDING SCIENCE, *supra* note 45, at 67-8. "Solid" here means "warranted;" but this does not lead to a vicious circle; eventually we reach experiential evidence, which neither has nor stands in need of warrant.

<sup>47</sup> See HAACK, DEFENDING SCIENCE, *supra* note 45, at 75-6 (providing a fuller argument why epistemic likelihoods do not satisfy the axioms of the mathematical calculus of probabilities). The thesis is not a new one. See JOHN MAYNARD KEYNES, A TREATISE ON PROBABILITY 28 (1921) (arguing that "[i]t is not even clear that we are always able to place [epistemic likelihoods] in an order of magnitude"); see also RICHARD VON MISES, PROBABILITY, STATISTICS AND TRUTH 97 (2nd rev. English ed. 1928) (arguing that "our probability theory has nothing to do with such questions as 'Is there a probability of Germany being involved in a war with Liberia?'").

warrant of the conclusion (as completing more of the crossword makes us more confident in the correctness of the completed entries if they all fit together, but undermines our confidence if it introduces inconsistencies). So a combination of pieces of evidence will warrant a conclusion to a higher degree than any of its components alone would do when, but only when, combining the various elements enhances supportiveness; enhances the independent security of evidence favorable to the conclusion; and/or enhances comprehensiveness by introducing further, no less supportive, elements.

\*

If we apply this rather abstract analysis to a schematic example based on the kinds of congeries of evidence typically encountered in toxic-tort cases, and look at the effect of combining evidence on supportiveness, independent security, and comprehensiveness, we will see how combining evidence can – as Justice Stevens and Judge Terry believed it could – enhance the degree of warrant of a causal conclusion.

Suppose the claim at issue is that exposure to substance S causes, or promotes, disorder D: e.g., that a pregnant woman's being exposed to Bendectin causes birth defects in her baby, or that exposure to PCBs promotes the development of lung cancer. The evidence, E, relevant to the conclusion, C, might include any or all of the following elements  $e_1, e_2, \dots, e_n$ :

- epidemiological evidence (from clinical trials or medical surveys) of the incidence of D among those exposed to S, as compared with its incidence among those not exposed to S;
- meta-analyses of such epidemiological studies, indicating what, if any, elevated risk of D is suggested by their combined data;
- evidence about whether the incidence of D in the population falls when S is withdrawn from the market (or cleared out of buildings, or whatever);
- evidence about what the components of S are (say a, b, and c), and of whether exposure to any other substance(s) containing one or more of these, or to chemicals of the same general type, is associated with elevated risk of D;

- evidence from *in vivo* studies indicating whether animals deliberately exposed to S develop D or precursors of D;
- evidence from *in vitro* studies indicating whether cells or embryos deliberately exposed to S develop D or precursors of D;
- evidence as to whether there is (are) any biological mechanism(s) by which exposure to S (or to a, b, and/or c) might cause D, or reasons for believing that S (or a, b, or c) could not cause D.

But the evidence with respect to a causal conclusion may also include a good deal of information of other kinds, bearing on it a bit less directly:

- meta-evidence with respect to all the types of evidence listed above: for example, evidence about what is required of a well-designed and well-executed epidemiological, toxicological, *in vitro*, or *in vivo* study (e.g., what variables need to be controlled for, etc.), and what constitutes a well-designed and well-conducted meta-analysis (e.g., what determines which studies are good enough to be included in a meta-analysis and which are best disregarded);
- background information about what other factors (such as genetic susceptibilities) might contribute to the development of D;
- background information (or conjecture) about what proportion of cases of D derive from what kinds of known (or suspected) cause;
- relevant chemical, biological, physiological, genetic, etc., theory potentially bearing on S or on D;<sup>48</sup>

---

<sup>48</sup> For example, as late as the early 1950s it was widely believed that nothing harmful could cross the placenta from mother to fetus. Since 1955, however, it was known that substances with a molecular weight of less than 1,000 *could* cross the placenta into fetal blood. ROCK BRYNNER and TRENT STEPHENS, DARK REMEDY: THE IMPACT OF THALIDOMIDE AND ITS REVIVAL AS A VITAL MEDICINE 12 (2001).

- ideas about what, in what is not yet known, is reasonably believed to be potentially relevant to the etiology of D.

And there may, additionally, be evidence (meta-meta-evidence?) about the sources of all these kinds of evidence,<sup>49</sup> bearing indirectly on its credibility, and hence, at one remove, on the credibility of C:

- evidence that relevant studies were published after peer review in well-respected journals, or were published by editorial privilege in low-status journals, or were not published at all;
- evidence about who conducted the relevant research: perhaps the manufacturer, or scientists funded by the manufacturer (and whether the research was paid for out of the manufacturer's research budget, or out of its litigation fund), or university scientists receiving some perks from the manufacturer, or independent scientists with no connections to either party in a suit;
- evidence that this witness is (or is not) a repeat testifier in such cases as this, that his resumé shows that he is (or is not) a professional expert witness rather than an active scientist; etc.
- evidence (meta-meta-meta-evidence?) as to whether studies funded by manufacturers tend to be more favorable to their products than studies conducted independently,<sup>50</sup> how often

---

<sup>49</sup> Because legal players are not experts in epidemiology, toxicology, etc., and don't have the kind of extensive background knowledge required to make judicious judgments of plausibility, this kind of (indirect, external) evidence probably plays a more significant role in legal contexts than, ideally, it might.

<sup>50</sup> In fact, many studies-of-studies confirm that company-funded research on drugs or medical devices is significantly more likely than independent research to be favorable to the company's products. See e.g., Richard A. Davidson, *Sources of Funding and Outcomes of Clinical Trials*, 1 J. GEN. INTERNAL MED. 155 (1986); Paula Rochon et al., *A Study of Manufacturer-Supported Trials of Non-Steroidal Anti-Inflammatory Drugs in the Treatment of Arthritis*, 154 ANNALS INTERNAL MED. 157 (1994); Lee S. Friedman and Elihu D. Richter, *Relationship Between Conflict of Interest and Research Results*, 19 J. INTERNAL. MED. 52 (2004). While legal commentators tend to be preoccupied with litigation-driven science, we should not forget that marketing-driven science may also be tendentious. See e.g. Kevin P. Hill et al., *The ADVANTAGE Seeding Trial: A Review of Internal Documents*, 149.4 ANNALS INTERNAL MED. 251, 251 (2008) (arguing that internal documents



peer reviewed papers are retracted,<sup>51</sup> whether papers in lower-ranked journals are retracted more often than those published in more prestigious fora, etc., etc.

E may be complete (i.e., include evidence of all the kinds listed), or it may be incomplete; and it may be all positive (i.e., supportive of C over not-C), or all negative, or mixed. For obvious reasons, in the cases that come to court the evidence is usually incomplete, mixed or, most often, both; for if it were entirely unambiguous one way or the other, either the claim would never have been brought, or it would have been settled.

No single element of a congeries of evidence such as E will be sufficient by itself to establish a causal conclusion. The effects of S on animals may be different from its effects on humans. The effects of b when combined with a and c may be different from its effects alone, or when combined with x and/or y.<sup>52</sup> Even an epidemiological study showing a strong association between exposure to S and elevated risk of D would be insufficient by itself: it might be poorly-designed and/or poorly-executed, for example (moreover, what constitutes a well-designed study – e.g., what controls are needed – itself depends on further information about the kinds of factor that might be relevant). And even an excellent epidemiological study may pick up, not a causal connection between S and D, but an underlying cause both of exposure to S and of D; or possibly reflect the fact that people in the very early stages of D develop a craving for S. Nor is evidence that the incidence of D fell after S was withdrawn sufficient by itself to establish causation – perhaps vigilance in reporting D was relaxed after S was withdrawn, or perhaps exposure to x, y, z was also reduced, and one or all of these cause D, etc.<sup>53</sup>

---

show that Merck's 1999 ADVANTAGE trial of Vioxx was "a seeding trial developed by Merck's marketing division to promote prescription of Vioxx (rofecoxib) when it became available ... in 1999.").

<sup>51</sup> The medical indexing service PubMed assigns a number, PMID (PubMed Identifier) to each article, and it is possible to search for e.g., "Retraction of Publication." On retractions of fraudulent work, see e.g., Laura Bonito, *The Aftermath of Scientific Fraud*, 124 CELL 873 (2006); Harold C. Sox and Drummond Rennie, *Research Misconduct, Retraction, and Cleansing the Medical Literature: Lessons from the Poehlman Case*, 144 ANNALS OF INTERNAL MEDICINE 609 (2006); Jennifer Couzin & Katherine Unger, *Cleaning Up the Paper Trail*, 312 SCIENCE 38 (2006).

<sup>52</sup> As, apparently, was the case with Thalidomide, which has been described as composed of "two rather innocuous compounds." See BRYNNER AND STEPHENS, DARK REMEDY, supra note 48, at 8 (quoting Dr. Robert Brent; on whom see note 53 *infra*).

<sup>53</sup> Dr. Robert Brent, the editor of TERATOLOGY, who testified repeatedly for Merrell Dow in the Bendectin cases as an expert on "secular trend data," emphasized that, although Bendectin had

But combining evidence, as in my schematic example, can help exclude explanations other than S's causing D, and thus warrant the conclusion more firmly. To understand under what conditions E would warrant C to a higher degree than any of  $e_1$ ,  $e_2$ , . . . ,  $e_n$  individually, we need to look at the effect of combining these on the overall supportiveness of E, on the independent security of each element of E, and on the comprehensiveness of E.

(i) *Supportiveness*: How supportive evidence is of a conclusion depends, to put it quite briefly and roughly, on how well the evidence and the conclusion fit together to form an explanatory account. So combined evidence will support a conclusion better than its component parts individually if the conjunction of E and C makes a better explanatory account than the conjunction of  $e_1$  and C, a better explanatory account than the conjunction of  $e_2$  and C, . . . , and so on. What makes the degree of support given to C by E greater than the degree of support given to C by  $e_1$ , the degree of support given to C by  $e_2$ , etc., is *how tightly its components interlock* to form an explanatory account. For example, evidence of a biological mechanism by which S might bring about D interlocks with epidemiological evidence of increased risk of D among those exposed to S to explain a formerly-unexplained aspect of the story; evidence that S contains b, and that it is b that is associated with increased risk of D, interlocks with epidemiological evidence of an increased risk of D among those exposed to S to make a formerly superficial explanation deeper; and background biological, physiological, chemical, etc., theory interlocks with evidence of the risks to humans of exposure to S to increase the scope of a formerly narrow explanatory account.

For the elements of E to interlock at all, the same terms ("S," "b," "D," etc) must occur throughout, as they do in my schematic list; and the elements will interlock more tightly the more narrowly these terms are characterized, i.e., the more specific they are. For example, joint support will be enhanced more if "D" is "small-cell lung cancer" than if it is simply "lung cancer" or "cancer," or if it is "limb-reduction birth defects"

---

been off the U.S. market since 1983, the rate of reported birth defects had remained steady; but by a parallel argument to the one in the text, this is insufficient by itself to rule out a casual conclusion. In fact, we know that after the withdrawal of Bendectin, some doctors began prescribing vitamin B6 and half a Unisom tablet for morning-sickness, and that doxylamine succinate, the suspect ingredient in Bendectin, is also an ingredient of Unisom (and of Nyquil). See Janelle Yates, *Nausea and Vomiting of Pregnancy: Q&A with T. Murphy Goodwin*, 16.8 OBG MGMT. 54, 55 (2004) (recommending vitamin B6 and, if vomiting continues, adding 12.5 mg doxylamine by halving the over-the-counter Unisom tablet). On Nyquil (as well as Unisom), see also Joseph Sanders, *From Science to Evidence: The Testimony on Causation in the Bendectin Cases*, 46 STANFORD L. REV. 1, 10 (1993).

rather than “birth defects”; if “b” is “doxylamine succinate” rather than “anti-histamine,” or “Benlate”<sup>54</sup> rather than “fungicide”; and so on.

The elements of E will also interlock more tightly the more physiologically similar the animals used in any animal studies are to human beings. The results of tests on hummingbirds or frogs would barely engage at all with epidemiological evidence of risk to humans, while the results of tests on mice, rats, guinea-pigs, or rabbits would interlock more tightly with such evidence, and the results of tests on primates more tightly yet. Of course, “similar” has to be understood as elliptical for “similar in the relevant respects;” and *which* respects are relevant may depend on, among other things, the mode of exposure: if humans are exposed to S by inhalation, for example, it matters whether the laboratory animals used have a similar rate of respiration. (Sometimes animal studies may themselves reveal relevant differences; for example, the rats on which Thalidomide was tested were immune to the sedative effect it had on humans; which should have raised suspicions that rats were a poor choice of experimental animal for this drug.)<sup>55</sup> Again, the results of animal tests will interlock more tightly with evidence of risk to humans the more similar the dose of S involved. (One weakness of Joiner’s expert testimony was that the animal studies relied on involved injecting massive doses of PCBs into a baby mouse’s peritoneum, whereas Mr. Joiner had been exposed to much smaller doses when the contaminated insulating oil splashed onto his skin and into his eyes.)<sup>56</sup> The timing of the exposure may also matter, e.g., when the claim at issue is that a pregnant woman’s being exposed to S causes this or that specific type of damage to the fetus.

Again, the elements of E will interlock more tightly the more closely *in vitro* studies match the conditions of human exposure. For example, the plaintiffs in *Castillo v.*

---

<sup>54</sup> In *Castillo v. du Pont*, Benlate was the fungicide to which Ms Castillo claimed she had been exposed, and which she believed had caused her baby’s birth defect, severely underdeveloped eyes (microphthalmia). *Castillo*, 854 So. 2d at 1264.

<sup>55</sup> BRYNNER AND STEPHENS, DARK REMEDY, *supra* note 48, at 48. “It was disturbing that humans responded to thalidomide by lapsing into a ‘deep, natural sleep,’ but rats did not . . . The fact that no lethal dose for rats could be found seemed doubly disturbing . . . the rats simply weren’t absorbing the medicine.” Richardson-Merrell was the U.S. distributor for Thalidomide. *Id.* at 39. The drug was originally sold as a sleeping pill. *Id.* at 14. Later, after the Australian physician Dr. William McBride discovered that it helped with morning sickness, it was prescribed for that purpose. *Id.* at 22. Subsequently, Dr. McBride became a hero for drawing attention to the dangers of Thalidomide, and then notorious, after he was found to have faked results in animal studies in an effort to draw attention to what he believed were the teratogenic effects of Bendectin. *Id.* at 27-29, 197-9.

<sup>56</sup> *Joiner*, 522 U.S. at 144.

*du Pont* go to great pains to show that the exposure of cells to Benlate in the *in vitro* studies to which they appealed was as nearly as possible the same as the exposure Ms. Castillo's unborn baby had allegedly undergone when his mother was accidentally sprayed with Benlate being used on neighboring fields.<sup>57</sup>

(ii) *Independent security*: combining evidence may also enhance independent security (as the fact that this crossword entry interlocks with others which in turn interlock with others, . . . , and so on, gives you more reason to think that it is correct). To be sure, adding evidence from animal studies won't make a flawed epidemiological study any less flawed, and adding evidence of a physiological mechanism won't make a sloppily-conducted *in vitro* study any more rigorous. (This seems to be the point Skrabanek and McCormick are making when they explain that the "faggot fallacy" is fallacious because "a bundle of insecure evidence remains insecure.")<sup>58</sup> However, if we add to only modestly secure epidemiological evidence of an elevated risk of D among those exposed to S, the further evidence that there is a biological mechanism by which S leads to D, this additional evidence enhances the security of the conclusion of the epidemiological study. (Similarly, if I add a column of numbers and reach the answer *n*, but am not sure my answer is right because I was interrupted in the middle of my calculation, asking someone else to check the arithmetic and finding that they get the same answer properly increases my confidence in the answer I got the first time – even though it doesn't alter the fact that I was interrupted).

(iii) *Comprehensiveness*: E is of course more comprehensive than any of its components alone; and this may enhance the degree of warrant of C (as completing a new entry in a crossword puzzle in a way compatible with the existing entries gives you reason to be more confident in them all). If, for example, we add to epidemiological evidence indicating an elevated risk of D among those exposed to S (*e*<sub>1</sub>), evidence about the chemical composition of S and the damaging physiological effects of its components (*e*<sub>2</sub>), and evidence of the biological mechanism by which exposure to S causes D (*e*<sub>3</sub>), this combined evidence will warrant the causal conclusion to a higher degree than any component part of this evidence standing alone. (Evidence of a statistical association of smoking and lung cancer<sup>59</sup> warrants a causal conclusion to a higher degree if it is

---

<sup>57</sup> See e.g. *Castillo*, 854 So. 2d at 1274 ("Dr. Howard considered what clothes Donna Castillo was wearing when she was exposed, and her height and weight to determine the amount of skin exposed, and used DuPont's data to calculate the amount of benomyl [the suspect ingredient in Benlate] that would have been absorbed and passed through her system.").

<sup>58</sup> Skrabanek and McCormick, *supra* note 29, at 35. See also A. R. Feinstein, *Scientific Standards in Epidemiologic Studies of the Menace of Everyday Life*, 242 SCIENCE 1257 (1988).

<sup>59</sup> Richard Doll and Austin Bradford Hill, *Smoking and Carcinoma of the Lung: Preliminary Report*, 2

combined with evidence of a causal mechanism; statistical evidence that women are more susceptible than men would warrant a causal conclusion to a higher degree if it is combined with evidence of the role of female hormones in speeding things up.)<sup>60</sup> However, the degree of warrant will go down, rather than up, if the additional evidence is negative, or even less positive, than the rest. If, for example, we add to evidence from animal studies indicating an elevated risk of D in those exposed to S ( $e_1$ ), evidence that an epidemiological study finds *no* elevated risk in humans ( $e_2$ ), the degree of warrant given C by this combined evidence will be lower, not higher.

\*

What I have offered is a theoretical analysis, an abstract characterization of the determinants of evidential quality – an analysis powerful enough, as we have seen, to show that combined evidence may indeed warrant a casual conclusion better than any of

---

(4682) BRITISH MED. J. 739 (Sep. 30, 1950) (PMID 14772469); M. L. Levin et al., *Cancer and Tobacco Smoking: A Preliminary Report*, 143.4 J. AM. MED. ASSOC. 336 (May 27, 1950) (PMID 15415261); C. A. Mill and M. M. Porter, *Tobacco Smoking Habits and Cancer of the Mouth and Respiratory System*, 10.9 CANCER RESEARCH 539 (Sep. 1950) (PMID 14772728); Schrek et al., *Tobacco Smoking as an Etiologic Factor in Disease. Part I: Cancer*, 10.1 CANCER RESEARCH 49 (Jan. 1950) (PMID 15398042); E. L. Wynder and E. A. Graham, *Tobacco Smoking as a Possible Etiologic Factor in Bronchiogenic Carcinoma: A Study of 684 Proved Cases*, 143.4 J. AM. MED. ASSOC. 329 (May 27, 1950) (PMID 15415260). These five studies published in 1950 are now seen as groundbreaking. By 1953, 13 more such studies had appeared.

<sup>60</sup> Michaela Kreuzer et al., *Hormonal Factors and Risk of Lung Cancer in Women?*, 32 INT'L J. OF EPIDEMIOLOGY 263 (2003) (suggesting exactly this). *But see also* Leno Thomas, et al., *Lung Cancer in Women: Emerging Differences in Epidemiology, Biology, and Therapy*, 120.1 CHEST 370, 370 (July 2005) (“[e]merging evidence suggests there are differences in the pathogenesis and possibly increased susceptibility to lung cancer in women”); International Early Lung Cancer Action Program Investigators, *Women’s Susceptibility to Tobacco Carcinogens and Survival After Diagnosis of Lung Cancer*, 290.2, J. AM. MED. ASSOC. 180, 180 (July 12, 2006) (“[w]omen appear to have increased susceptibility to tobacco carcinogens but have a lower rate of fatal outcome of lung cancer compared to men”); Geoffrey C. Kabat et al., *Reproductive and Hormonal Factors and Risk of Lung Cancer in Women: A Prospective Cohort Study*, 120 INT. J. CANCER 2214, 2214 (2007) (“[s]everal lines of evidence suggest that endocrine factors may play a role in the development of lung cancer in women, but the evidence is limited and inconsistent”); Diana C. Marqu ez-Garb an et al., *Estrogen Receptor Signaling Pathways in Human Non-Small Cell Lung Cancer*, 72 STEROIDS 135, 136 (2007) (“[e]strogen status appears to be a significant factor in lung cancer in women ...”); Patricia O’Keefe and Jyoti Patel, *Women and Lung Cancer*, 24.1 SEMINARS IN ONCOLOGY NURSING 3, 4 (Feb., 2008) (“[w]omen may be more susceptible to the carcinogenic effects of lung carcinogens than men. ... Research in this area is ongoing and is highly debated”); Neal D. Freedman, et al., *Cigarette Smoking and Subsequent Risk of Lung Cancer in Men and Women: Analysis of a Prospective Cohort Study*, 9 THE LANCET 649 (Jul. 2008), available at <http://oncology.thelancet.com> (last visited Sep. 29, 2008) (suggesting that the claim that women are more susceptible than men is questionable).

its components. It does not, however, purport to be a decision-procedure for arriving at a conclusion about the reliability or otherwise of causation (or other) evidence. Nevertheless, it sheds some light on the kerfuffle over “weight of evidence methodology” in *Joiner*. It should already be apparent that G.E.’s accusation that Joiner’s experts have committed a fallacy<sup>61</sup> in supposing that combined evidence warrants their causal conclusion better than its individual elements rests on a mistake. But it should also be clear – though it is, perhaps, not quite so obvious – that Joiner’s appeal to “weight of evidence methodology” is itself a bit misleading, at least if it is intended to suggest that there is anything like an algorithm or protocol, some effective, mechanical procedure for calculating the combined worth of evidence.

This is also apparent if one looks closely at the 1986 EPA Guidelines for Carcinogen Risk Assessment<sup>62</sup> to which Joiner’s attorneys refer.<sup>63</sup> These guidelines advise that “[t]he question of how likely an agent is to be a human carcinogen should be answered in the framework of a weight-of-the-evidence judgment”;<sup>64</sup> however, they don’t use the phrase “weight of evidence methodology,” or offer anything like an algorithm for determining the joint weight of evidence. The section headed “Categorization of Overall Weight of Evidence for Human Carcinogenicity” simply describes how categories are assigned: “(1) The weight of evidence in human studies or animal studies is summarized; (2) these lines of information are combined to yield a tentative assignment to a category (see Table 1); (3) all relevant supportive information is evaluated to see if the designation of the overall weight of evidence needs to be modified”; which amounts to little more than “we look at all the available evidence and use our judgment to assess what it shows.” Table 1 – described as “for illustrative purposes” only – is a little more specific: for example, it indicates that a substance is categorized as a human carcinogen only when there is “sufficient” epidemiological evidence, and as a probable human carcinogen if there is “limited” epidemiological evidence but “sufficient” evidence from animal studies.<sup>65</sup> But this amounts to little more than requiring epidemiological evidence before putting a substance in the highest-risk category – provided that this epidemiological evidence is “sufficient.”

The more recent, 2005 EPA Guidelines include a section with the curious but revealing heading “Weight of Evidence Narrative,” which explains that the EPA still

---

<sup>61</sup> See *supra* note 29 and accompanying text.

<sup>62</sup> Environmental Protection Agency, Guidelines for Carcinogen Risk Assessment, 51.186 Federal Register 3392, 34000 (Sep. 24, 1986).

<sup>63</sup> *Joiner*, 522 U.S. 136 (Stevens, J. dissenting, citing Brief for Defendants, 4041).

<sup>64</sup> EPA Guidelines for Carcinogen Risk Assessment (1986), *supra* note 62, at 33996.

<sup>65</sup> *Id.* at 34000.

“emphasizes the importance of weighing all of the evidence in reaching conclusions about the human carcinogenic potential of agents” but, moving away from the “step-wise approach” of the 1986 guidelines, now takes “a single integrative step.” Data from epidemiological studies are generally preferred, “but all of the [epidemiological, *in vivo*, *in vitro*, toxicological, etc.] information . . . could provide valuable insights.”<sup>66</sup> So far, perhaps, not much more helpful than the 1986 guidelines; but as one reads on, there are several observations worth noting. First, these guidelines use the same metaphor of “fitting together” that I have, quite independently, used here:

[t]he narrative explains the kinds of evidence available and *how they fit together* in drawing conclusions, and . . . points out significant issues/strengths/limitations of the data and conclusions.<sup>67</sup>

Second, they take for granted – just as I have here, in articulating *to what degree* evidence warrant a conclusion, and when a congeries of evidence warrants a conclusion *to a higher degree* than its components – that warrant is a matter of degree:

descriptors [“human carcinogen,” “probable human carcinogen,” etc.] represent points along *a continuum of evidence*, . . . there are gradations and borderline cases . . .<sup>68</sup>

And third, they acknowledge the distinction I have stressed between frequency probabilities (as in “the probability that a randomly selected Swede is a Protestant is *n*%,” or “the probability that a 60-year old American male will live to be 75 is *m*%”) and epistemic likelihoods (as in “it is overwhelmingly likely that PCBs are carcinogenic”):

[a]lthough the term ‘likely’ can have a probabilistic connotation in other contexts, its use as a weight of evidence descriptor *does not correspond to a quantifiable probability* of whether the chemical is carcinogenic.<sup>69</sup>

But when it comes to the core question, “what determines the weight of evidence?”, these guidelines fall back on the so-called “Bradford Hill criteria,” drawn from Sir Austin Bradford Hill’s now-classic 1965 lecture, *The Environment and Disease*.<sup>70</sup>

---

<sup>66</sup> Environmental Protection Agency, Guidelines for Carcinogen Risk Assessment, EPA/630P-03/001F (March 2005), 1-11.

<sup>67</sup> *Id.* at 1-12.

<sup>68</sup> *Id.* at 2-51 (emphasis added).

<sup>69</sup> *Id.* at 2-53 (emphasis added).

<sup>70</sup> Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 PROCEEDINGS

These, however, are not criteria for determining the quality of evidence generally but, as Hill's title suggests, are focused specifically on medical causation evidence (especially evidence relating to occupational exposure); moreover, they apply only in a situation where there is already statistical evidence of an elevated risk of D among those exposed to S. What Bradford Hill offers is a list of nine aspects of a known association between S and D that should be considered in arriving at a conclusion as to whether or not the connection is causal:

- (1) *Strength*: i.e., how large the increase of risk of D is in those exposed to S;
- (2) *Consistency*: i.e., whether the association between S and D been observed by different persons, in different places and times, and under different circumstances;
- (3) *Specificity*: i.e., whether the association is specifically between this substance or occupational exposure, and this disease;
- (4) *Temporality*: i.e., whether exposure to S precedes D (rather than, e.g., being associated with the early stages of D);
- (5) *Biological gradient*: i.e., whether the incidence of D rises as exposure to S rises;
- (6) *Plausibility*:<sup>71</sup> i.e., whether the causal hypothesis fits with current biological knowledge;
- (7) *Coherence*: i.e., interpreting the data as showing causation should not ("seriously") conflict with known facts about the history and biology of

---

OF THE ROYAL SOCIETY OF MEDICINE 205 (1965). According to the FEDERAL REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 376 (Federal Judicial Center, 2nd ed. 2000) Bradford Hill was amplifying criteria proposed by the U.S. Surgeon General in assessing the relationship between smoking and lung cancer. U.S. Dept. of Health, Educ., and Welfare, Smoking and Health Report of the Advisory Committee of the Surgeon General (1964).

<sup>71</sup> While I was writing this, the press reported that a new study finding Vytorin no more effective than a placebo with respect to a certain heart-valve condition, had also found an increased risk of cancer in those taking the drug; but that researchers "declared the latter finding 'implausible' and probably the result of chance." Their reason, I take it, was that there was no known mechanism that could plausibly be supposed to explain such a connection. Ron Winslow and Shirley S. Wang, *More Vytorin Bad News Hits Merck, Schering*, WALL ST. J., Jul. 22, 2008, at B1, B2.



the disease;

(8) *Experiment*: i.e., whether the incidence of D falls if preventive action is taken to reduce exposure to S;

(9) *Analogy*: i.e., whether there is some similarity to other known cases of a causal connection.

It is worth noting that Bradford Hill acknowledges that “[n]one of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis, and none can be required as a *sine qua non*.”<sup>72</sup>

It is hardly surprising that these “Bradford Hill criteria” have proved so durable,<sup>73</sup> for they contain much good sense. But they are not really “criteria,” at least as that term is sometimes understood; not, that is, a decision-procedure, or even a checklist that could be followed mechanically. (The legal term “indicia” might be less misleading.) Evidence may, for example, satisfy some of these and not others, or may satisfy some in high degree and others in lower degree; and Bradford Hill says nothing about how to assess success on one of these indicia against failure on that, or how to compare hypotheses one of which does well on this and poorly on that, and the other poorly on this and well on that. This is hardly surprising, either. For in fact – as my theoretical account suggests, and as the EPA’s curious word “narrative” hints – assessing the worth of complex evidence is, in a sense, inevitably a matter of judgment; that is to say, someone experienced in the field may see that the causal claim is (to use Bradford Hill’s word) “plausible,” because he has brought to bear a whole mesh of background knowledge and presumption – a mesh of background knowledge which, however, he doesn’t, and perhaps couldn’t, articulate fully. Indeed, it is precisely because the assessment of complex evidence is a matter of judgment in this sense that even well-qualified and highly-competent experts may reasonably disagree; for unless and until the evidence is overwhelming one way or the other, subtle differences in the unarticulated complex of background knowledge and presumption each scientist brings to the table may produce different assessments.

Still, it may be helpful to map Bradford Hill’s somewhat unsystematic list of

---

<sup>72</sup> Bradford Hill, *The Environment and Disease*, *supra* note 70, at 299.

<sup>73</sup> They appear, for example, not only in the 2005 EPA guidelines, *supra* note 66; but also in SANDERS, BENEDICTIN ON TRIAL, *supra* note 15, at 55-6, in the FEDERAL REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 70, at 375-6; and in Kenneth J. Rothman and Sander Greenland, eds., MODERN EPIDEMIOLOGY 24-28 (2nd ed. 1998).

“criteria” or “indicia” onto the more articulated structure of the account I have proposed. “Consistency” amounts in effect to acknowledgment that combined evidence from different sources, provided it all points in the same direction, improves the warrant of the conclusion that the connection is causal. Bradford Hill’s “coherence,” “biological plausibility” and “analogy” seem to correspond to the kinds of evidence included in my list under “background knowledge” – whatever is known about a potential mechanism or mechanism by which S might cause D, any biological, physiological, etc., theory, with which the causal conclusion would fit, and so on. “Specificity” corresponds to the connection I make between tightness of fit of the elements of E and how narrowly “S” and “D” are specified; and “temporality” to the fact that an association between S and D found in even an excellent epidemiological study could be the result of a common cause of exposure to S and of D, or of the fact that the presence of D itself leads to exposure to S. “Experiment” corresponds to what I have described as evidence about whether the incidence of D is changed when exposure to S is deliberately reduced (“secular trend data,” as Dr. Brent calls it); and “biological gradient” is reflected at least in my part by my observations about the extent and the manner of exposure to S.

While the “strength” of the association between S and D, i.e., how large the increased risk is, finds its place in my account as relevant to ruling out the possibility that an apparently elevated risk is the result of chance, it is the *first* thing on Bradford Hill’s list. He appeals to the example of the incidence of scrotal cancer in chimney-sweeps – among whom, he reports, even as late as the 1920s the death rate from scrotal cancer was *200 times* the rate among those not exposed to tar or mineral oils.<sup>74</sup> It is worth noting that the connection with “tar or mineral oils” is already built into Bradford Hill’s example, and that the possibility that men in the early stages of scrotal cancer are somehow thereby attracted to chimney-sweeping as a profession seems so remote as to be negligible. But the main reason this factor is less prominent in my schematic example than in Bradford Hill’s list is, simply, that the kinds of case that come to court will surely not be those where the association is *so* strong that the inference to a causal conclusion is virtually guaranteed, but are far more likely to be those where, after a drug tested even in large clinical trials goes on the market, vastly more people take it, and there is evidence, or suspicion, that there may be unanticipated risks to some sub-group of the

---

<sup>74</sup> Bradford Hill, *The Environment and Disease*, *supra* note 70, at 295, *citing* Richard Doll, *Cancer*, in L. J. Witts, ed., *MEDICAL SURVEYS AND CLINICAL TRIALS: SOME METHODS AND APPLICATIONS OF GROUP RESEARCH IN MEDICINE* 333 (2nd ed. 1964). (According to Dr. Doll, in 1775 Percivall Pott reported that “cancer of the scrotum was characteristically a disease of chimney sweeps;” and in 1933 J. W. Cook at al. proved that “3:4-benzpyrene was responsible for the carcinogenic action of pitch on the skin of animals.” *Id.* at 333 *citing* J. W. Cook et al., *The Isolation of a Cancer-Producing Hydrocarbon from Coal Tar*, 1 J. CHEM. SOC. 395 (1933).).

population.

### 3. Answering Some Contested Questions

The theoretical apparatus now in place suggests (at least the beginnings of) answers to a range of epistemological questions that have often bedeviled toxic-tort litigation – questions about proof of general causation (the main topic here), and even some questions about proof of specific causation.

*Is epidemiological evidence of an elevated risk of D among those exposed to S essential to proof of general causation?*<sup>75</sup> “Epidemiologic studies,” the 1986 EPA guidelines observe, “provide unique information about humans who have been exposed to suspect carcinogens.”<sup>76</sup> “[D]escriptive” epidemiological studies, they continue, “are useful in generating hypotheses and providing supportive data,” but “can rarely be used to make a causal inference”; however, “analytical” case-control or cohort studies “are especially useful in assessing risks to exposed humans.”<sup>77</sup> Obviously, well-designed and well-conducted epidemiological studies showing an elevated risk would significantly increase the degree of warrant of a causal conclusion; and, of course, unlike animal studies, where there is always a question whether the animals used are enough like human beings in the relevant respects, epidemiological studies involve human subjects (which, no doubt, is why Table

---

<sup>75</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 721 F. Supp. 570, 575 (S.D. Cal. 1989) (holding that, given that there was a vast body of epidemiological evidence regarding Bendectin, expert opinion not based on epidemiological evidence was not admissible). *See also, e.g.*, *Grimes v. Hoffman-LaRoche, Inc.*, 907 F. Supp. 33, 35 (D.N.H. 1995) (excluding Dr. Lerman’s testimony that Acutane played a role in Mr. Grimes’ developing cataracts in part on the grounds that “[r]ather than relying on epidemiological data, Dr. Lerman bases his general causation opinion primarily on scientific theory, an in vitro experiment, and what he considers certain ‘generally accepted’ scientific facts”); *Sutera v. The Perrier Group of America*, 986 F. Supp. 655 (D. Mass. 1997) (excluding plaintiffs’ expert testimony because they have “produced no scientific peer-reviewed epidemiological studies which would associate APL [acute promyelocytic leukemia] ... and benzene exposure” at the relevant levels); *In re Rezulin Products Liability Litigation*, 369 F. Supp. 2d 398, 411 (S.D.N.Y. 2005) (excluding plaintiffs’ expert testimony that the diabetes drug Rezulin caused “silent” liver damage, in part on the grounds that “[t]here are no clinical trials and no observational epidemiological studies supporting the plaintiffs’ position”); *In re Bextra and Celebrex Marketing Sales Practices and Product Liability Litigation*, 524 F. Supp. 2d 1166, 1175 (N.D. Cal. 2007) (excluding plaintiffs’ expert testimony that Celebrex could cause cardiovascular effects at a dose of 200 mg. daily in part on the grounds that “there are no randomized controlled trials or meta-analyses of such trials or meta-analyses of observational studies that find an association between Celebrex 200 mg/d and risk of heart attack or stroke”).

<sup>76</sup> EPA Guidelines for Carcinogen Risk Assessment (1986), *supra* note 62, at 33995.

<sup>77</sup> *Id.*

1 in the 1986 EPA guidelines effectively allows epidemiological studies to trump animal studies). Nevertheless, if there is sufficient positive evidence of other kinds, a causal conclusion might be warranted to a non-negligible degree even in the absence of epidemiological evidence.

This is particularly significant when, for one reason or another, no relevant epidemiological studies are available, or possible.<sup>78</sup> Michael Gottesman argues that “it is quite rare” that “conclusive human epidemiological evidence is available”;<sup>79</sup> for when it is suspected that a drug or chemical may be harmful, manufacturers are likely either to institute “protective procedures for future use of the product” or else to withdraw it from the market, which makes such epidemiological work much more difficult. For example, he continues, PCBs had been routinely used in electrical transformers until reports began to link them to certain cancers, and they were banned in 1977;<sup>80</sup> after that, they were no longer used in transformers, and there was no longer any realistic possibility of conducting epidemiological studies of a possible link between PCBs and the kind of cancer Mr. Joiner developed.<sup>81</sup>

In any case, it is important to be clear that “there is no epidemiological evidence of an elevated risk of D in those exposed to S” is not equivalent to “there is epidemiological evidence that there is no elevated risk of D among those exposed to S.” (Unlike the so-called “faggot fallacy,” confusing these two very different propositions really *is* a fallacy.) For example, early on there was no evidence one way or the other about whether patients who took Vioxx for less than 18 months had elevated cardiovascular risk – and early on in the Vioxx litigation, Merck argued as if this were evidence that there was no elevated risk among patients who took the drug only for a short time.<sup>82</sup> But when later studies looked at short-term Vioxx use, they found evidence

---

<sup>78</sup> See *Castillo*, 854 So. 2d. at 1269-70 (reporting that the plaintiff’s expert argued that “clinical epidemiological studies are not available because Benlate is a toxic chemical and thus not suitable for human experiment,” and that “in cases where exposure is very rare to begin with, there are inherent problems with epidemiological studies because a scientist cannot [ethically] expose a human to a known teratogen in order to study the effects.”).

<sup>79</sup> Michael H. Gottesman, *From Barefoot to Daubert to Joiner: Triple Play or Double Error?*, 48 ARIZ. L. REV. 753, 767 (1998) (Mr. Gottesman argued *Daubert* and *Joiner* for the plaintiffs at the Supreme Court).

<sup>80</sup> 90 Stat. 2003, 2025 (1976), 15 U.S.C. § 605 (e).

<sup>81</sup> Gottesman, *supra* note 76, at 767.

<sup>82</sup> “In an admission that could undermine one of its core defenses in Vioxx-related lawsuits, Merck said yesterday that it had erred when it reported in early 2005 that a crucial statistical test showed that Vioxx caused heart problems only after 18 months of use.” Alex Berenson, *Merck Admits a Data Error on Vioxx*, N. Y. TIMES, May 3, 2006, at C1, available at 2006 WLNR 9291555.

suggesting that the risk went up as early as the first dose.<sup>83</sup> This brings home the lesson: that the absence of evidence that *p* is just that – an absence of evidence; it is not evidence that not-*p*.

*If there are relevant epidemiological studies, and they find no elevated risk of D among those exposed to S, is this always and inevitably fatal to a claim of general causation?* No, not always or necessarily. If they are good studies, yes; but if they are significantly flawed in ways that makes it likely that they underestimated the risk, their negative results are not fatal to such a claim. In *Blum v. Merrell Dow*, for example, defendant's expert Dr. Shapiro acknowledged under cross-examination that his epidemiological study had lumped together women who took Bendectin during the period of pregnancy in which fetal limbs were forming, and women who took the drug only after the limbs had formed, and so may have underestimated any elevated risk of limb-reduction defects.<sup>84</sup> Or, to take a more recent example, we now know that the VIGOR study, Merck's first large clinical trial of Vioxx, kept track of the gastrointestinal effects of Vioxx for longer than it kept track of the cardiovascular effects; and as a result, failed to find a statistically significant elevated risk of heart attack and stroke.<sup>85</sup>

---

In *Plunkett v. Merck & Co. (In re. Vioxx Products Liability Litigation)*, 410 F. Supp. 2d 565, 596-7 (E.D. La. 2005) the plaintiffs moved to exclude Merck's testimony that Vioxx only causes prothrombotic effects if taken for 18 months or more; but was denied on the grounds that both parties relied on the same study (the APPROVe study), while the court should be concerned only with methodology, not with the conclusions drawn.

<sup>83</sup> Patricia McGettigan and David Henry, *Cardiovascular Risk and Inhibition of Cyclooxygenase: A Systematic Review of the Observational Studies of Selective and Non-Selective Inhibitors on Cyclooxygenase*, 296.2 JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION 1633 (Oct. 4, 2006) citing studies finding elevated risk with early Vioxx use: W. A. Ray et al., *Cyclo-oxygenase-2 Selective Non-Steroidal Anti-Inflammatory Drugs and Risk of Serious Coronary Heart Disease*, 360 LANCET 1071 (2002); D. H. Solomon et al., *Relationship Between Selective Cyclooxygenase-2 Inhibitors and Acute Myocardial Infarction in Older Adults*, 109 CIRCULATION 2068 (2004); Linda Levesque et al., *Time Variations in the Risk of Myocardial Infarction Among Elderly Users of Cox-2 Inhibitors*, published electronically at www.cmj.ca (May 2, 2006) and, abridged, in 174.11 CANADIAN MEDICAL ASSOCIATION JOURNAL 1563 (May 23, 2006).

<sup>84</sup> *Blum v. Merrell Dow*, 33 Phila. Co. Rptr. 193, 215-7 (Ct. Comm. Pleas. Pa. 1996). See also Susan Haack, *What's Wrong With Litigation-Driven Science? An Essay in Legal Epistemology*, 38-3 SETON HALL L. REV. 1053, 1066-69 (2008) (scrutinizing the defendant's expert testimony in *Blum*).

<sup>85</sup> See David Armstrong, *Bitter Pill: How the New England Journal of Medicine Missed Warning Signs on Vioxx – Medical Weekly Waited Years to Report Flaws in Article that Praised Pain Drug – Merck Seen as "Punching Bag,"* WALL ST. J., May 15, 2006, at A1 (exploring problems with Merck's reporting of the results of the VIGOR study, see also Haack, *supra* note 25, at 804-7 (discussing problems in the VIGOR study)).

*Is it acceptable to disregard, or simply and on principle to exclude, epidemiological studies the results of which are not statistically significant?*<sup>86</sup> No. To be sure, the less statistically robust a study, the less it contributes to the warrant of a causal conclusion. But the crucial point is that statistical significance is a matter of degree, and that the cut-off degree conventionally accepted is just that, a convention – a cut-off point adopted by the relevant scientific community, and set high to ensure that the risk of false positives is minimized.<sup>87</sup> Bradford Hill was right when he wrote, almost half a century ago, that the then fast-growing emphasis on statistical significance meant that “too often . . . we grasp the shadow and lose the substance” as we “deduce ‘no significance’ from ‘no statistical significance.’”<sup>88</sup> But the trend he deplored is now firmly-entrenched practice.<sup>89</sup> And unfortunately, as Rothman et al. observe in their amicus brief in *Daubert*, “a factfinder who is told that a body of data is not ‘statistically significant’ is made to believe that the data has no value”; and, as they continue, the “talismanic” phrase “statistically significant” can create the completely misleading impression that statistically significant data are infallible.<sup>90</sup>

*Dunn v. Sandoz Pharmaceutical Corporation*<sup>91</sup> is especially fascinating for its confusion over this question. In brief: Ms. Dunn had sued Sandoz, the manufacturers, alleging that their anti-lactation drug, Parlodel, had caused her post-partum stroke; but the court excluded her general causation expert, Dr. Kulig, on the grounds that his testimony was insufficiently reliable to pass muster under *Daubert*.<sup>92</sup> Dr. Kulig testified:

---

<sup>86</sup> See *In re. Bextra*, *supra* note 75 (excluding plaintiffs’ testimony on the risk of adverse events in those taking 200 mg. of Celebrex a day in part on the grounds that the epidemiological studies found no statistically significant association); see also *Daubert*, 727 F. Supp. at 570 (giving this as part of the reason for excluding the Dauberts’ expert testimony).

<sup>87</sup> Moreover, there are different ways of calculating statistical significance, which sometimes give different results, and the choice of which is sometimes itself a matter of controversy. See e.g. Keith J. Winstein, *Boston Scientific Stent Study Flawed*, WALL ST. J., Aug. 14, 2008, at B6 (reporting such a controversy).

<sup>88</sup> Hill, *supra* note 70, at 299-300.

<sup>89</sup> Winstein, *supra* note 87, at B1 (noting that “medical journals typically won’t publish” studies the results of which are not statistically significant).

<sup>90</sup> Brief of Professor Kenneth Rothman et al., as Amici Curiae in Support of Petitioners, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 570 (1993) (No. 92-102) at \*4.

<sup>91</sup> *Dunn v. Sandoz Pharmaceuticals Corporation*, 275 F. Supp. 2d 672, 677, 680 (M.D.N.C. 2003).

<sup>92</sup> *Dunn*, 275 F. Supp. 2d at 681; see also *Soldo v. Sandoz Pharm. Corp.*, 241 F. Supp. 2d 434 (W.D. Pa. 2003); *Caraker v. Sandoz Pharm. Corp.*, 172 F. Supp. 2d 1046 (S.D. Ill. 2001). Dr. Kulig had also proposed to testify to the same effect in these cases, but had been excluded. Ms. Caraker’s attorneys, by the way, had likened the expert evidence they offered as *fitting together like the pieces of a jigsaw puzzle* to establish causation. *Caraker*, 172 F. Supp. 2d. at 1048. (Before I adopted the crossword analogy, I had for a while worked with Michael Polanyi’s analogy, likening

“I believe causation exists because I’ve applied the Bradford Hill criteria”;<sup>93</sup> but the court agrees with Sandoz that Dr. Kulig had *mis*applied those criteria by failing to notice that they kick in only when there is already epidemiological evidence of an association between a substance and a disorder. So far, fair enough. But then the court quietly slips in an additional phrase: Dr. Kulig would have needed “to have [had] a *statistically significant* study as the beginning point for the application of the Bradford Hill criteria.”<sup>94</sup> The court may have been correct in suspecting that Dr. Kulig’s application of the Bradford Hill criteria was largely decorative,<sup>95</sup> and was certainly correct in pointing out that these criteria presuppose some evidence of an association as their starting point; but evidently it was *not* aware of Bradford Hill’s skeptical attitude to the insistence on statistical significance.

*Is it appropriate to disregard (or in principle to exclude) evidence from animal studies?*<sup>96</sup> Of

---

the work of science to putting together a huge jigsaw. See MICHAEL POLANYI, *THE REPUBLIC OF SCIENCE: ITS POLITICAL AND ECONOMIC THEORY* (1962); reprinted in Marjorie Grene, ed., *KNOWING AND BEING* 49, 51-2 (1969).

<sup>93</sup> *Dunn*, 275 F. Supp. 2d at 677.

<sup>94</sup> *Dunn*, 275 F. Supp. 2d at 680.

<sup>95</sup> In any case, as I argued above, the Bradford Hill “criteria” can be at best *indicia* of a causal connection. (Dr. Kulig’s testimony suggests a certain ambivalence on this point: “[t]he toxicologic community, my peers, use Bradford-Hill extensively. ... [In my testimony] *I’ve taken the extra step and applied a published, generally accepted criteria to the analysis...* And the Bradford-Hill criteria, in my opinion, it’s a generally accepted scientific methodology to the analysis of adverse drug reactions;” however, he also acknowledges that “you may interpret the evidence differently.”) *Dunn*, *supra* note 91, at 677-8. Reading his affidavit in this case, I notice that he writes as if Bradford Hill had provided a check-list, running through it commenting, *e.g.*, “this criteria [temporality] is clearly met,” etc.; whereas Bradford Hill seems well aware both that many of his criteria can be met in varying degrees, and that it is necessary to use one’s judgment in deciding how likely it is that the relationship is causal. (I also note, for the record, that “criteria” is plural, not, as Dr. Kulig seems to think, singular.) Affidavit of Kenneth Kulig, M.D., FAACT, FACMT at ¶ 27-30, *Dunn v. Sandoz*, 275 F. Supp. 2d 672 (M.D.N.C. 2003) (No. 1:98 CV 00912), 2000 WL 34616176.

<sup>96</sup> See *e.g.*, *Metabolife Int’l v. Wornick*, 72 F. Supp. 2d 1160, 1169 (S.D. Cal. 1999) (excluding Metabolife’s scientific evidence, in part on the grounds that as a matter of law animal studies are inadmissible, “due to the uncertainties in extrapolating from effects on mice and rats to humans.”). In 2001 the U.S. Court of Appeals for the 9th Circuit reversed this exclusion. See *Metabolife Int’l v. Wornick*, 264 F.3d 832, 842-43 (9th Cir. 2001) (holding that the District Court abused its discretion in excluding the animal studies); see also *In re Silicone Gel Breast Implants Prod. Liab. Litig.*, 318 F. Supp. 2d 879, 891 (C.D. Cal. 2001) (excluding plaintiffs’ evidence from animal studies on the grounds that “[e]xtrapolations of animal studies to human beings are generally not considered reliable in the absence of a scientific explanation of why such extrapolation is warranted.”) (quoting *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1410 (D. Or. 1996)). In *Joiner*, the District Court had agreed with G.E. that the animal studies on which

course not. Obviously such studies can contribute to the warrant of a causal conclusion – the more so, the better-designed and better-conducted they are, using appropriate animals, doses, modes of delivery, times of delivery, etc. Of course, and no less obviously, there is always the possibility that animals are adversely affected by S, but humans are not, and vice-versa;<sup>97</sup> and if well-designed and well-conducted tests on animals show an elevated risk of D with exposure to S, but well-designed and well-conducted epidemiological studies show no elevated risk of D in humans exposed to S, we would rightly suspect that there might be relevant physiological differences of which we are not yet aware.

*Is epidemiological evidence of at least a doubling of risk (epistemologically) essential to establishing specific causation?* – i.e., to go beyond the general claim that exposure to S sometimes causes or promotes D, to the specific claim that it was his or her exposure to S that caused or promoted this plaintiff's D, is it necessary to show that exposure to S doubles the risk of D? This is the requirement under e.g., New Jersey law, and was imposed by Judge Kozinski when he reheard *Daubert* on remand from the Supreme Court.<sup>98</sup> But it rests on a confusion. The idea behind such a requirement is, presumably, that only if exposure to S at least doubles the risk of D can we infer that the odds are

---

his experts relied were inadequate to establish that Joiner's exposure to PCBs had promoted his cancer; at appeal, Joiner's attorneys (unwisely) argued as if the issue was whether animal studies, as such, can ever be a proper foundation for an expert's opinion. *See* Gen. Elec. Co. v. Joiner, 522 U.S. 136, 144 (1997).

<sup>97</sup> “[O]ne can usually rely on the fact that a compound causing an effect in one mammalian species will cause it in another species. This is a basic principle of toxicology . . . ” FEDERAL REFERENCE MANUAL ON SCIENTIFIC TESTIMONY, *supra* note 70, at 410. However, animal studies have two disadvantages: the difficulty in extrapolating to humans because “differences in absorption, metabolism, and other factors may result in interspecies variation in responses”; and because “the high doses customarily used in animal studies” leave open questions about dose-response relation in humans. *Id.* at 346.

<sup>98</sup> Rehearing *Daubert* on remand from the Supreme Court, Judge Kozinski argued that the Dauberts' experts would have to be excluded under the new standards, as they had been under the *Frye* Rule; finding that, unless an expert claimed to show that Bendectin at least doubled the risk of birth defects, he would have to be excluded on grounds of irrelevance. *See Daubert*, 43 F.3d at 1320-1321 (“California tort law requires plaintiffs to show not merely that Bendectin increased the likelihood of injury, but that it more likely than not caused *their* injuries”) (*citing* Jones v. Ortho Pharm. Corp., 163 Cal. App. 3d 396 (1985)). The court continued: “[i]n terms of *statistical* proof, this means that the plaintiffs must establish ... that [their mothers' taking Bendectin] more than doubled” the risk. *See Daubert*, 43 F.3d at 1320; *see also id.* at 1321 (*citing* DeLuca v. Merrell Dow Pharm. Inc., 911 F.2d 941, 958 (3rd Cir. 1990), where the requirement of New Jersey law that plaintiffs must show that more likely than not Bendectin caused Amy DeLuca's birth defects is interpreted as meaning that “the relative risk of limb reduction defects arising from the epidemiological data ... will, at a minimum, have to exceed ‘2’”).



that the plaintiff, who was exposed to S and developed D, developed D *because* he or she was exposed to S. But this idea rests on a confusion of statistical probabilities with epistemic likelihoods; and it is clear on reflection that a doubling of statistical risk is neither necessary nor sufficient for proof of specific causation.

Epidemiological evidence of a doubling of risk is *not sufficient* for specific causation: first, because if the study showing a doubling of risk is poorly-designed or poorly-executed, we would have only *a low epistemological likelihood of a greater than 50% statistical probability*; and second, because even a well-designed and well-conducted study might also show that those subjects who develop D when exposed to S have some characteristic in common – older patients rather than younger, perhaps, or women rather than men, or the sedentary rather than the active – and our plaintiff might be an elderly, sedentary female. And epidemiological evidence of a doubling of risk is *not necessary* for specific causation, either: first, because studies that fail to show a doubling of risk may be flawed – for example, by failing to take account of the period of pregnancy in which subjects are exposed to S, or by failing to take account of the fact that subjects are included who may have been exposed to S in cold medication or sleep-aids;<sup>99</sup> and second, because even a good epidemiological study indicating to a high degree of epistemic likelihood that there is a doubling of risk may also indicate that those subjects who develop D have some characteristic (such as being over 50 or sedentary or subject to allergies or whatever) that this plaintiff lacks.<sup>100</sup>

There is a related problem with another argument sometimes encountered, that since (say), it is believed on reliable evidence that 10% of cases of D are genetic, and 20% caused by environmental factors, while the causes of the remaining 70% are unknown, the odds are that this plaintiff's D was not caused, as alleged, by exposure to S. But here the confusion between statistical and epistemic probabilities is overlaid by confusions of two other kinds: a false presumption that the cause of D must be *either* genetic *or* environmental (when there may be interaction between the two); and treating

---

<sup>99</sup> See *supra* note 53 and accompanying text.

<sup>100</sup> In a footnote, Judge Kozinski acknowledges this problem, at least in part: “[n]o doubt, there will be unjust results with this standard. If a drug increases the likelihood of birth defects, but doesn’t more than double it, some plaintiffs whose injuries are attributable to the drug will be unable to recover”; but dismisses it with the comment that “[t]here is a converse unfairness under a regime that allows recovery to everyone who may have been affected by the drug” and that this is a matter to be sorted out by the states. *Daubert*, 43 F.3d at 1320 n.13. He also acknowledges the possibility that we might have evidence that a plaintiff belongs to a more-than-usually susceptible sub-class of the population, but notes that the plaintiffs in *Daubert* had offered no evidence to this effect. See *id.* at 1321 n.16.

“unknown” as if it referred to another type of cause, like “genetic” or “environmental” – when really, obviously, it is an expression of ignorance. If a plaintiff argues that it was exposure to S that caused him to develop D, and the defendant replies that this is unlikely, since we know that 70% of cases of D stem from unknown causes, the defendant’s response is defective – because if the plaintiff’s claim is true, what we *think* we know about what proportion of cases of D are caused by known factors, and what by unknown factors, may not, after all, be genuine knowledge.

When Donald Rumsfeld made that notorious remark about “unknown unknowns,”<sup>101</sup> the topic, of course, was Iraqi intelligence. Perhaps I was the only person in the country who didn’t laugh derisively; at any rate, from a strictly epistemological perspective, at least, Secretary Rumsfeld had a genuinely important point: not only may we not have all the evidence we know would be relevant (the “known unknowns” in Rumsfeldese); there may be evidence we don’t have that we don’t even realize is relevant. This – the Rumsfeld Problem of unknown unknowns – is also relevant to the next question on my list.

*Can we infer from the fact that the causes of D are as yet unknown, and that a plaintiff developed D after being exposed to S, that it was this exposure that caused Ms. X’s or Mr. Y’s D?*<sup>102</sup> No. Such evidence would certainly give us reason to look into the possibility that S is the, or a, cause of D. But loose talk of “inference to the best explanation” disguises the fact that what presently seems like the most plausible explanation may not really be so – indeed, may not really be an explanation at all. We may not know all the potential causes of D, or even which other candidate-explanations we would be wise to investigate.

---

<sup>101</sup> Donald H. Rumsfeld, *U.S. Sec’y of Def., Dept. of Def., News Briefing*, Feb. 12, 2002, available at <http://www.defenselink.mil.transcripts/transcript.aspx?transcriptid=2636> (last visited Oct. 15, 2008). (“Reports that say that something hasn’t happened are always interesting to me. Because as we know, there are known knowns, there are things that we know we know. We also know that there are known unknowns, that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know.”)

<sup>102</sup> See e.g. *Rosen v. Ciba-Geigy Corp.*, 78 F.3d 316, 318 (7th Cir. 1996) (holding that the district court had not abused its discretion in excluding Dr. Fozzard’s testimony that Mr. Rosen’s heart attack was caused by his having worn a nicotine patch for three days before it occurred: “[w]hen an unusual event follows closely on the heels of another unusual event, the ordinary person infers a causal relation . . . . But lay speculations on medical causality, however plausible, are a perilous basis for inferring causation . . . .”)

#### 4. The Legal Argument

Under *Daubert* courts must screen proffered expert testimony<sup>103</sup> for relevance and (“evidentiary”) reliability. It is worth pausing for a moment to point out that relevance, like reliability, is a factual matter. Whether (and to what degree) p is relevant to q, that is, is not a matter of pure logic, but depends on facts about the world: if, but only if, astrology is true, for example, the position of the planets at the time of your birth is relevant to how things will turn out for you this week. But the focus in this paper is on the reliability prong.

Reliability, I take it, is a matter of degree; admissibility, by contrast, is categorical: a witness is either allowed to testify, or to testify to this or that question,<sup>104</sup> or not. So a court determining whether or not testimony is admissible is normally<sup>105</sup> imposing a sharp, yes-or-no dichotomy on a continuum of degrees of reliability.<sup>106</sup> The mismatch between the categorical nature of admissibility and the gradational character of reliability has been even more marked since 2000, when FRE 702 was revised to require that expert testimony be based on “sufficient” data, “reliably” arrived at and “reliably” applied to facts at issue.<sup>107</sup> And the fact that a party facing a *Daubert* challenge to their proffered expert testimony must show “by a preponderance of the evidence” that this testimony meets the legal standard of reliability compounds the complexities. What they must show, apparently, is that *it is more likely than not* that this testimony is

---

<sup>103</sup> Some might prefer to put this a little differently: that *Daubert* clearly imposed this requirement with respect to scientific testimony, but only when the Supreme Court clarified the scope of *Daubert* in *Kumho Tire* was it clear that the requirement also applies to expert testimony other than the scientific. *Kumho Tire v. Carmichael*, 525 U.S. 137 (1999).

<sup>104</sup> See e.g. *U.S. v. Llera Plaza*, Nos. CR 98-362-10, 98-362-11, 98-362-12 (E.D. Pa. Jan 7 2002). Judge Pollack ruled that while fingerprint examiners’ testimony was admissible on certain matters, “the parties will not be permitted to present testimony expressing an opinion of an expert witness that a particular latent print matches, or does not match, the rolled prints of a particular person and hence is, or is not, the fingerprint of that person.” *Id.* at 19.

<sup>105</sup> *But see* Transcript of Bench Ruling at 1484, *U.S. v. Brown*, N. 05 Cr. 538 (JSR) (S.D.N.Y. June 18, 2008) (reasoning that admissibility under *Daubert* need not be construed as categorical, and permitting ballistics examiners to testify only that their conclusions were more likely than not; and observing that the court “had a discussion about a year ago with Prof. Dan Capra [of Columbia and Fordham Law Schools] and asked him “was Rule 702 supposed to be an absolute rule, in the sense of either it is in or it is out” and he said no, not at all”). See also *U.S. v. Glynn*, No. 06 Cr. 580 (JSR), 2008 WL 4293317, at \*1 (S.D.N.Y. Sept. 22, 2008) (referring to the court’s ruling in *Brown*).

<sup>106</sup> See Dale Nance, *Two Concepts of Reliability*, (APA) NEWSLETTER ON PHILOSOPHY AND LAW, Fall 2003 at 123.

<sup>107</sup> FED. R. EVID. 702.

*likely enough* to satisfy the reliability prong of *Daubert*.” Well: I have worked in epistemology for many years now, but I have to say that it’s a mystery to me what this means.

But the problem most immediately relevant to present purposes is that *Daubert* seems to impose a kind of evidentiary atomism<sup>108</sup> that pulls against the more holistic character of most causation evidence. The problem is very noticeable in *Joiner*, when the Supreme Court looked one by one at (some of) the studies Joiner’s experts’ would have cited had they been admitted, and finds that none of them would pass muster under *Daubert*. But Judge Kozinski’s 1995 ruling on remand reveals that the problem derives from *Daubert* itself. Because the law had changed since the trial court granted summary judgment for Merrell Dow in 1989, Judge Kozinski argued, there might be a case for allowing the plaintiffs the opportunity to make a showing that their proffered expert testimony met the new standard;<sup>109</sup> however, he went on, there was no point in doing this if it was already clear that their experts would have to be excluded under *Daubert*, as they had been under *Frye*. And in fact, he continued, this *was* already clear. Looking at each of the *Daubert*’s experts’ proffered testimony one by one, Judge Kozinski observes first that all but one of them proposed only to testify that there was a possibility that Bendectin causes birth defects, and didn’t even claim, let alone show, that a mother’s taking the drug more than doubled the risk, and so would have to be excluded under the relevance prong;<sup>110</sup> and then that Dr. Palmer the only expert who claimed more, that Bendectin caused Jason *Daubert*’s birth defects, simply *had no* methodology to speak of, and so would have to be excluded under the reliability prong.

And this atomistic strategy is implicit in the *Daubert* Court’s ruling, according to which *each item* of expert evidence is to be screened for (relevance and) reliability. To be admissible,  $e_1$  must be (relevant and) reliable,  $e_2$  must be (relevant and) reliable,  $e_3$  must be (relevant and) reliable, . . . , and so on.<sup>111</sup> But if my epistemological argument is correct, the combination of  $e_1, e_2, \dots, e_n$  may warrant a causal conclusion better than any

---

<sup>108</sup> Also sometimes called “corpuscularism.” See Thomas McGarity, *Our Science is Sound Science and Their Science is Junk Science: Science-Based Strategies for Avoiding Accountability and Responsibility for Risk-Producing Products and Activities*, 52.4 U. KAN. L. REV. 897, 921 (2004).

<sup>109</sup> See *Daubert*, 43 F.3d at 1315. Indeed, the rhetoric of *Daubert* 1993 was that the new standard was *more* hospitable to the admission of expert testimony than the old, “austere” *Frye* Rule. See *Daubert*, 509 U.S. at 589 (“that austere standard, absent from, and incompatible with, the Federal Rules of Evidence, should not be applied in federal trials”).

<sup>110</sup> See also *supra* note 98 and accompanying text.

<sup>111</sup> See McGarity, *supra* note 108, at 924 (“[u]nder the corpuscular approach, a study is either valid or it is invalid, and it is either relevant or irrelevant. A conclusion based on invalid or irrelevant studies cannot be relevant or reliable and must therefore be rejected”).

of its components alone – may be, in *Daubert's* terminology, more reliable than any of its components.

It might be thought – for a while I thought myself – that this difficulty could be avoided if *Daubert* were interpreted as requiring, not that each item of expert testimony reliably enough indicate the ultimate conclusion that exposure to S causes or promotes D, but that each item reliably enough indicate the conclusion of the study referred to: e.g., that the data from an epidemiological study reliably enough indicate the conclusion “there is an elevated risk of n%, among those exposed to S, of developing D,” that the data from an animal study reliably enough indicate the conclusion “when animals of this kind are exposed to this dose of S, delivered in this manner, m% of them develop D,” . . . and so on. But while there is arguably some justification for this interpretation of the ruling in Justice Blackmun’s footnote about the intended meaning of evidentiary reliability,<sup>112</sup> it does not, I’m afraid, solve the problem.

“[D]elusiveness,” Oliver Wendell Holmes once shrewdly observed, “is the source of fallacy throughout the law.”<sup>113</sup> And indeed, it is not clear that giving a precise meaning to “preponderance of the evidence” would be desirable, even if it were possible. But for the purposes of my argument it doesn’t matter what, exactly, the “preponderance of the evidence” standard – a phrase which, interestingly enough, has the “weighing” metaphor built in<sup>114</sup> – amounts to. For the essential point is that, *however* one sets that standard, there could be instances in which the evidence is equally balanced, i.e., in which the evidence warrants C and not-C to the same degree; and in such circumstances even a minimal increment of warrant one way or the other would give a “preponderance” in favor of C, or against it. And while it is true that evidence  $e_i$ , favorable to C, will improve the warrant of C less if it is itself less than solid (and evidence  $e_k$ , unfavorable to C, will decrease the warrant of C less if it is itself less than solid), even such evidence might tip the scales, i.e., make the difference between “evenly balanced” and “marginally favors C over not-C,” or vice versa. And so, if some element of evidence that might have tipped the scale is excluded under the reliability prong of *Daubert*, this may actually impede assessment of the reliability of the scientific testimony in its entirety – because the jury will never hear any element that the court excludes on the grounds that it is insufficient by itself to meet the standard.

---

<sup>112</sup> *Daubert*, 509 U.S. at 590, n.9 (characterizing “evidentiary reliability”).

<sup>113</sup> *Truax v. Corrigan*, 257 U.S. 312, 342 (1921) (Holmes, J., dissenting).

<sup>114</sup> WEBSTER’S NINTH NEW COLLEGIATE DICTIONARY 929 (Merriman-Webster, 1991) (defining “preponderance” as “superiority in weight, power, importance, or strength”).

Of course, though factual truth is undeniably important to substantive justice, some rules of evidence – spousal privilege, for example, or FRE 407(b), under which evidence of subsequent repair is inadmissible – deliberately allow considerations of policy to preclude the presentation of evidence that might be highly relevant to the truth of facts at issue. Whether such policy-oriented rules are justifiable is a whole other issue, which I can't pursue here;<sup>115</sup> but in any case, FRE 702 is not such a rule, but is focused precisely and unambiguously on reliability.

More relevant to the present argument is the thought that courts excluding scientific testimony under the reliability prong of *Daubert* may (at least sometimes) be motivated by concern that a jury presented with a lot of weak evidence may draw an unwarranted conclusion. A jury may, indeed, be misled in this way: for it doesn't follow from the fact that, as I have argued here, a combination of pieces of evidence each individually insufficient may jointly warrant a conclusion to a higher degree than any component element, that *any and every* combination of such evidence warrants the conclusion to the required standard of proof. But a court may also be misled, perhaps in the opposite direction; for it doesn't follow from the fact that a combination of pieces of evidence each individually insufficient may also be jointly insufficient, that *any and every* combination of such evidence fails to warrant the conclusion to the required degree. As this reveals, the root of the problem is that, while the legal system relies increasingly on scientific testimony, neither judges nor juries – nor attorneys, for that matter – are well-equipped to make judgments on scientific questions where even highly-qualified and competent experts may honestly, and reasonably, disagree.<sup>116</sup>

---

<sup>115</sup> *But see* Susan Haack, *Epistemology Legalized: Or, Truth, Justice, and the American Way*, 49 AM. J. JURIS. 43, 56-60 (2004) (briefly addressing this issue).

<sup>116</sup> My thanks to Pamela Lucken, of the University of Miami Law Library, for capable research assistance; to Lee Tilson for information about retractions of medical articles and Celeste Monforton for references on smoking and lung cancer; and to Mark Migotti, Adina Schwartz, Marina Teplitsky, and, especially, Joseph Sanders for helpful comments and suggestions.