

ARTICLES

THE HEURISTICS OF INTELLECTUAL DUE PROCESS: A PRIMER FOR TRIERS OF SCIENCE

ERICA BEECHER-MONAS*

Scientific evidence is an inescapable facet of modern litigation. The Supreme Court, beginning with the seminal case of Daubert v. Merrell Dow Pharmaceuticals, Inc., and continuing with General Electric Co. v. Joiner and Kumho Tire Co. v. Carmichael, has instructed federal judges to evaluate the scientific validity of such evidence in determining the evidence's admissibility. In this Article, Professor Erica Beecher-Monas argues that many judges ignore the science component of their "gatekeeping" duties, focusing instead on rules of convenience that have little scientific justification. As a result, she demonstrates that judges reject even scientifically uncontroversial evidence that would have little trouble finding admissibility under the pre-Daubert "general consensus" standard and admit evidence that is scientifically baseless. Such faulty analysis of scientific evidence deprives litigants of intellectual due process from judges and undercuts the proper functioning and credibility of the judicial system. Beecher-Monas contends that understanding certain basic principles underlying all fields of science will enable judges to make better admissibility decisions. Based on the language of science and criteria scientists use to assess validity, as well as the Supreme Court's requirements in Daubert, Joiner, and Kumho Tire, Beecher-Monas proposes a five-step framework for sound analysis of scientific evidence. She then demonstrates the usefulness of the heuristic in two cases where applying the heuristic would have changed the outcome dramatically. The framework proposed in this Article will allow triers of science to make scientifically justifiable admissibility assessments, and in so doing will give litigants in cases involving scientific evidence the intellectual due process they deserve.

Introduction	1564
I. Evaluating Scientific Argument: What Is the Scientific Method Anyway?	1572
A. Scientific Method: The Myth	1573
B. Scientific Method: The Reality	1578
1. Probabilistic Reasoning	1580
2. Falsifiability, Criticism, and Rationality	1584
3. Explanatory Power	1588

* Assistant Professor of Law, University of Arkansas at Little Rock. B.A., 1976, M.S. (Anatomy), 1978, J.D., 1988, University of Miami; LL.M., 1995, Columbia University. This Article is written in partial fulfillment of the J.S.D. requirements of the Columbia University School of Law. The author is deeply grateful to the late Curtis J. Berger for his comments and encouragement, and to Frank P. Grad, Martha A. Fineman, Theresa Beiner, and John DiPippa for their helpful comments and criticism. For her excellent research assistance and unflagging enthusiasm, the author thanks Patti Stanley.

II. The Heuristic: A Guide for Judges in Deciding the Admissibility of Scientific Evidence.....	1589
A. What's the Theory? Examine the Explanatory Power of Theory and Hypothesis	1590
B. Examine All the Available Data.....	1591
C. Use Assumptions to Fill the Gaps	1596
1. Statistics and Numbers Count	1599
2. The Type of Study Counts	1604
3. Is Individualization a Valid Assumption?	1625
D. Examine the Methodology.....	1626
E. Probabilistic Assessment of Expert Conclusions: Putting It All Together	1631
III. Applying the Heuristic.....	1636
A. <i>Wright v. Willamette Industries, Inc.</i>	1637
1. What's the Theory?.....	1637
2. Examine All the Available Evidence	1638
3. Use Valid Assumptions to Fill the Gaps	1640
4. Examine the Methodology.....	1644
5. Probabilistic Assessment of the Link Between Data and Hypothesis	1644
B. <i>State v. Council</i>	1647
1. What's the Theory?.....	1649
2. Examine All the Available Evidence	1650
3. Use Valid Assumptions to Fill the Gaps	1652
4. Examine the Methodology.....	1652
5. Probabilistic Assessment of the Link Between Data and Hypothesis	1654
Conclusion	1656

INTRODUCTION

Presenting scientific evidence in court once seemed like a fairly straightforward proposition. Like most litigators, I used to believe my job was merely to find the right expert, to ask the right questions, and then to get out of the way of the testimony.¹ Now, finding the right expert, crucial though it may be, is only the first step in a complex process that involves educating the court about the methods and ter-

¹ For over 70 years, the admissibility of scientific expert testimony depended only on whether the evidence purported to reflect a consensus of the relevant scientific community. This "general acceptance" standard was articulated in *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), which explained that "while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." *Id.* at 1014.

minology of science. Ever since the Supreme Court, in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,² instructed the federal judiciary to make admissibility determinations based on an analysis of the scientific validity of the proffered testimony, and on whether the testimony “fits” the issues in the case, there has been a major paradigm shift in the way courts and litigants approach scientific evidence.³ And, just in case anyone thought *Daubert* was an aberration that the federal courts could soon forget, the Supreme Court reiterated the trial judge’s mandate to review testimony for scientific validity and “fit” in

² 509 U.S. 579 (1993).

³ *Daubert* involved birth defect claims relating to the antinausea drug Bendectin. Plaintiffs proffered expert causation testimony, which the trial and appellate courts had ruled inadmissible. See *Daubert v. Merrell Dow Pharms., Inc.*, 727 F. Supp. 570, 575-76 (S.D. Cal. 1989), aff’d, 951 F.2d 1128, 1131 (9th Cir. 1991), vacated and remanded, 509 U.S. 579, 597-98 (1993). The Supreme Court vacated and remanded, categorically dispatching the general acceptance test used by the lower courts in their determinations, finding it to be an “austere standard, absent from, and incompatible with, the Federal Rules of Evidence.” *Daubert*, 509 U.S. at 589. The *Daubert* Court announced the new standard for scientific evidence in light of its reading of Federal Rule of Evidence 702, which provides: “If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.” Fed. R. Evid. 702. The Court ruled that judges must examine proffered expert testimony to determine “whether the reasoning or methodology underlying the testimony is scientifically valid and . . . whether that reasoning or methodology properly can be applied to the facts in issue.” *Daubert*, 509 U.S. at 592-93.

Daubert is a federal standard, involving construction of the Federal Rules of Evidence. Yet most states have adopted rules modeled on the Federal Rules. See Paul Giannelli & Edward Imwinkelried, *Scientific Evidence: The Fallout from the U.S. Supreme Court’s Decision in Kumho Tires*, *Crim. Just.*, Winter 2000, at 12, 15 (analyzing impact of *Daubert* on states and noting that “numerous jurisdictions have rejected *Frye* in favor of the *Daubert* approach, other courts have opted to retain the *Frye* test”); see also Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 *Temp. L. Rev.* 55, 78-82 (1998). While adopting rules modeled on the Federal Rules does not necessitate that states adopt the construction given the Rules by the federal courts, many states have done so. See *id.* Even those states such as New York that continue to follow the general acceptance standard frequently feel compelled to enter the discourse about scientific validity encouraged by *Daubert*. See, e.g., *People v. Wernick*, 674 N.E.2d 322, 324 (N.Y. 1996) (“No threshold evidentiary foundation whatsoever was offered that acknowledged the *validity* or existence of defense counsel’s postulate to warrant these experts using this kind of extrapolated material to bolster their expert opinions.” (emphasis added)); *Clemente v. Blumenberg*, 705 N.Y.S.2d 792, 798-99 (Sup. Ct. 1999) (noting that although New York does not follow *Daubert*, New York law contains many principles inherent in *Daubert*, including importance of judge’s role as gatekeeper in determining scientific validity of expert evidence); *Wahl v. American Honda Motor Co.*, 693 N.Y.S.2d 875, 877-78 (Sup. Ct. 1999) (using *Daubert* standard to admit engineering testimony since *Frye* test in New York only applies to “novel” or scientific evidence); *Collins v. Welch*, 678 N.Y.S.2d 444, 446-47 (Sup. Ct. 1998) (engaging in validity analysis while acknowledging New York’s use of *Frye* general acceptance standard). For a further discussion of *Daubert*, its interpretation of Rule 702, and its impact on state and federal law, see Beecher-Monas, *supra*, at 78-82.

General Electric Co. v. Joiner,⁴ and extended its scope to technical as well as scientific evidence in *Kumho Tire Co. v. Carmichael*.⁵

Like all paradigm shifts, this one has created a great deal of angst among those who operated under the old regime,⁶ in which counting scientific noses was enough of an inquiry to determine if proffered testimony met minimal admissibility standards.⁷ In an earlier article, I explained that although *Daubert* provided some guidelines, its suggestions for resolving whether contested expert testimony meets the criteria of good science were sketchy, at best.⁸ In that article, I outlined

⁴ 522 U.S. 136 (1997). In *Joiner*, the Supreme Court upheld the traditional abuse of discretion standard of review for evidentiary issues, applying it even to scientific testimony. See *id.* at 146.

⁵ 526 U.S. 137 (1999). Before *Kumho Tire*, some courts had been avoiding the *Daubert* inquiry by classifying evidence such as psychological testimony, engineering testimony, and identification evidence such as handwriting analysis as “technical” or “soft science” rather than as “scientific.” See, e.g., *Jenson v. Eveleth Taconite Co.*, 130 F.3d 1287, 1297 (8th Cir. 1997) (noting that “[t]here is some question as to whether the *Daubert* analysis should be applied at all to ‘soft’ sciences such as psychology”); *United States v. DiDomenico*, 985 F.2d 1159, 1171 (2d Cir. 1993) (“‘[S]oft science’ expertise is less likely to overwhelm the common sense of the average juror than ‘hard science’ expertise” (quoting Cathleen C. Herasimchuk, A Practical Guide to the Admissibility of Novel Expert Evidence in Criminal Trials Under Federal Rule 702, 22 St. Mary’s L.J. 181, 198 (1990))); *United States v. Starzeczpyzel*, 880 F. Supp. 1027, 1041 (S.D.N.Y. 1995) (admitting handwriting analysis testimony, concluding that while such testimony would probably not pass *Daubert* scrutiny, since it is technical rather than scientific, it falls “outside the scope of *Daubert*”).

In *Kumho Tire*, the Supreme Court emphasized that the trial judge’s gatekeeping obligation applies to all expert testimony. See *Kumho Tire*, 526 U.S. at 147. The Court emphasized the importance of observing professional standards of intellectual rigor in all expert testimony, and concluded that “a trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony.” *Id.* at 152. The Court acknowledged that not all of the *Daubert* factors apply to all testimony, even to testimony based on traditional laboratory sciences, but emphasized that empirical support, the existence of protocols, and sound methodology are basic requirements for any expert testimony. See *id.* at 150, 153-58. For a further discussion of the impact of the Supreme Court’s trio of scientific evidence cases—*Daubert*, *Joiner*, and *Kumho Tire*—on scientific testimony, see Erica Beecher-Monas & Edgar Garcia-Rill, *The Law and the Brain: Judging Scientific Evidence of Intent*, 1 J. Appellate Prac. & Process 243, 246-51 (1999). For further commentary on the applicability of *Daubert* and this Article to “nonscientific” expert testimony, see *infra* note 32.

⁶ See, for example, *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1316 (9th Cir. 1995), where Judge Kozinski, on remand, complained that “though we are largely untrained in science and certainly no match for any of the witnesses whose testimony we are reviewing, it is our responsibility to determine whether those experts’ proposed testimony amounts to ‘scientific knowledge,’ constitutes ‘good science,’ and was ‘derived by the scientific method.’” For a critique of courts’ post-*Daubert* handling of scientific evidence, see *Developments in the Law: Confronting the New Challenges of Scientific Evidence*, 108 Harv. L. Rev. 1481, 1513-16 (1995).

⁷ See *supra* note 1 and accompanying text.

⁸ See Erica Beecher-Monas, *A Ray of Light for Judges Blinded by Science: Triers of Science and Intellectual Due Process*, 33 Ga. L. Rev. 1047, 1051 (1999) (“In *Daubert*, the Supreme Court’s explanation . . . was cursory at best.”). The Court established four guide-

and critiqued various frameworks available to help federal judges in these decisions.⁹ These attempts at guidance, while important first steps, focus on the concerns of scientists in creating good science rather than on evaluating the unavoidably imperfect science that exists, and therefore do not resolve the problems judges face in making *Daubert* admissibility decisions.¹⁰ Understanding how experiments

lines for the judge to consider: (1) whether the proffered testimony is testable, (2) whether it has been subjected to peer review and publication, (3) whether protocols and error rates exist for the methodology, and (4) whether it has achieved general acceptance. See *Daubert*, 509 U.S. at 593-94. The "*Daubert* question" ordinarily arises in pretrial motions under Federal Rule of Evidence 104. These hearings, in which the trial judge hears and rules on proffers of scientific evidence, are becoming more frequent. Indeed, these hearings are commonly referred to as "*Daubert* hearings." For a discussion of the procedural aspects of the "*Daubert* question," see Beecher-Monas, *supra* note 3, at 76-78.

⁹ See Beecher-Monas, *supra* note 8, at 1076-93. For example, the Federal Judicial Center's manual for evaluating scientific evidence outlines a variety of key scientific points that may arise in admissibility determinations. See generally Federal Judicial Ctr., *Reference Manual on Scientific Evidence* (1994) [hereinafter *Federal Judicial Ctr. Manual*]. The post-*Daubert* legal literature on scientific evidence admissibility is extensive. See Thomas Lyons, *Frye, Daubert and Where Do We Go From Here?*, R.I. B.J., Jan. 1997, at 5, 25 (observing that *Daubert* "has spawned an enormous reaction" with average of over four articles per month appearing in legal periodicals analyzing decision and issues it raises). Courses have even sprung up to help familiarize judges with scientific issues. See, e.g., John M. Conley & David W. Peterson, *The Science of Gatekeeping: The Federal Judicial Center's New Reference Manual on Scientific Evidence*, 74 N.C. L. Rev. 1183, 1205 n.139 (1996) (referring to authors' courses for judges).

¹⁰ For example, shortly after *Daubert*, Bert Black and coauthors offered a list of nine guidelines to help judges evaluate good science. See Bert Black et al., *Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge*, 72 Tex. L. Rev. 715, 782-85 (1994). These nine factors are (1) explanatory power, (2) falsifiability, (3) logical consistency, (4) scope of testing, (5) consistency with accepted theories, (6) subsequent application and use by the scientific community, (7) precision, (8) post-hypothesis testing, and (9) peer review and publication. See *id.* Never intended to be used as a checklist, the explicit purpose of the guidelines was to explain the concerns of scientists in doing their work. See *id.* at 782.

One useful framework that has been largely ignored by the judiciary is that provided by the Environmental Protection Agency (EPA) for its scientific validity determinations. See Environmental Protection Agency Proposed Guidelines for Carcinogen Risk Assessment, 61 Fed. Reg. 17,960, 17,961 (1996) [hereinafter *EPA Proposed Guidelines*] (emphasizing importance of considering all evidence in "addressing the likelihood of human carcinogenic effects of the agent and the conditions under which such effects may be expressed"). Agency administrators for regulatory agencies such as the EPA routinely evaluate competing scientific claims in their permitting decisions and risk assessments. The context of EPA decisionmaking is far more analogous to *Daubert* judicial validity determinations than the context discussed in previous studies of creating the best possible empirical tests of proposed hypotheses. Now that *Joiner* has confirmed the standard of review for scientific evidence, both trial court admissibility determinations and agency determinations such as the EPA's are subject to the same "abuse of discretion" standard. See *General Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997) (using "abuse of discretion" standard of review for trial court scientific evidence admissibility determinations); see also 5 U.S.C. § 706(2)(A) (1994) (using "arbitrary, capricious . . . abuse of discretion, or otherwise not in accordance with law" standard of review for administrative agency actions). As in evidentiary determinations, the basis for agency review is whether the facts and logic used in

should be designed is unquestionably useful to judges in evaluating scientific testimony. But it is not enough. Judges are not in the business of designing experiments to minimize flaws in concept and execution. They are in the business of determining whether the flawed studies—and all studies are flawed in some respects—that underlie the expert's opinion really support the expert's testimony. Thus, in the absence of concrete guidance, judges and lawyers remain perplexed about the required inquiry.

In this Article, I seek to provide a rough, but useful, guide for evaluating scientific testimony, based on a synthesis of the Supreme Court's evidentiary trilogy, the Environmental Protection Agency's (EPA) evidentiary framework,¹¹ and teachings from the philosophy of science.¹² Judges, after all, are not the only outsiders who must evaluate scientific evidence. Scientists who work outside of a given field critique each others' work all the time by taking information gleaned from one discipline and applying it to another; this is how science advances.¹³ How is this possible without intimate knowledge of the particular type of research being discussed? Common understandings about probabilistic and analogy-based reasoning underlie all scientific disciplines.¹⁴ Even nonscientists can learn this kind of reasoning.¹⁵

arriving at the determinations are supportable. See, e.g., *Natural Resources Defense Council, Inc. v. EPA*, 902 F.2d 962, 968 (D.C. Cir. 1990) (explaining that court must "carefully review the record to ascertain that the agency has made a reasoned decision based on 'reasonable extrapolations from some reliable evidence'" (quoting *Natural Resources Defense Council v. Thomas*, 805 F.2d 410, 432 (D.C. Cir. 1986))), vacated in part, 921 F.2d 326 (D.C. Cir. 1991).

¹¹ See *supra* note 10.

¹² I draw heavily from the work of Karl Popper, Imre Lakatos, Ernest Nagel, Patrick Suppes, and others. Although these philosophers often disagree with each other, they share many areas of common understanding, and each offers insights about particular aspects of the validity determination.

¹³ See Gero Decher, *Fuzzy Nanoassemblies: Toward Layered Polymeric Multicomposites*, 277 *Science* 1232, 1232 (1997) (observing how "materials science has developed into an interdisciplinary field" that draws from "organic, polymeric, and even biological" disciplines); Sheila Jasanoff et al., *Conversations with the Community: AAAS at the Millennium*, 278 *Science* 2066, 2066 (1997) (describing interdisciplinary research as "basic to the conduct of science"); Jeffrey Mervis, *NSF Revamps Graduate Training Grants*, 275 *Science* 918 (1997) ("It has become a truism that science is increasingly multidisciplinary . . .").

¹⁴ Definitions of probabilistic thinking are elusive. See Theodore Colton, *Statistics in Medicine* 63 (1974) ("There is no completely satisfactory definition of probability."). What I mean by probabilistic thinking, and the notion on which scientists generally agree, is the frequentist definition: The "probability of an event is the event's *long-run relative frequency* in repeated trials under similar conditions." *Id.* (emphasis added). Long-run relative frequency refers to the proportion of times that a particular trait or event being measured occurs within a given population. See Martin Sternstein, *Statistics* 15 (1994). "In other words, the *probability* of an event A is its relative frequency of occurrence—or the proportion of times the event occurs—in a large number of trials repeated under virtually identical conditions." Marcello Pagano & Kimberlee Gauvreau, *Principles of Bios-*

The object of demystifying scientific argument¹⁶ and making it more accessible to lawyers and judges is not to transform lawyers and judges into amateur scientists,¹⁷ but to help them resolve a legal policy issue: whether a scientific hypothesis proffered by experts is useful in resolving a legal dispute. The purpose of the admissibility inquiry, after all, is not to decide whose expert is correct, but whether the expert can provide information to help the factfinder resolve an issue in the case.¹⁸ That is, the judge must decide whether a descriptive claim about the world has sufficient indicia of reliability and relevance to the case at hand to enter the courtroom. For scientific evidence, the question is whether the testimony has met the standards and methods

tistics 127 (2d ed. 2000). Thus, it is an error, for example, to think that one can know the probability of disease causation after exposure to a toxic chemical in any given case. See Sander Greenland & James M. Robins, *Epidemiology, Justice, and the Probability of Causation*, 40 *Jurimetrics* 321, 328 (2000) (explaining that “when an exposure is known to be harmful in *some cases*, available data from epidemiology and biology are simply incapable of telling us whether a given case was ‘more probably than not’ harmed by exposure” (emphasis added)). Probabilistic thinking recognizes that the most that can be said in this example is that exposure caused a certain excess incidence of disease in a particular population under certain conditions. For further discussion of probabilistic reasoning, see *infra* Part I.B.1.

¹⁵ “Though the details of science may be remote from common experience, nonscientists can understand the fundamental characteristics that separate valid science from pale imitations.” Black et al., *supra* note 10, at 720; see also Francisco J. Ayala & Bert Black, *Science and the Courts*, 81 *Am. Sci.* 230, 238 (1993) (arguing that judges have ability to analyze scientific testimony and apply scientific standards to assess its validity).

¹⁶ An argument is “[a]ny inferential analysis leading to a stated conclusion.” 2 David A. Schum, *Evidence and Inference for the Intelligence Analyst* 15 (1987). Scientific inference drawing consists of proffering empirical data to support or refute hypotheses explaining observations about the natural world. See Wesley C. Salmon, *The Foundations of Scientific Inference* 1-53 (1967) (summarizing arguments about falsifiability, testability, and confirmation).

¹⁷ Chief Justice Rehnquist expressed this concern in his opinion in *Daubert*, where he worried that the majority was forcing judges “to become amateur scientists in order to perform [their] role.” *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 601 (1993) (Rehnquist, C.J., concurring in part and dissenting in part).

¹⁸ See Anthony Z. Roisman, *The Courts, Daubert, and Environmental Torts: Gatekeepers or Auditors?*, 14 *Pace Env'tl. L. Rev.* 545, 548 (1997) (observing that “when experts offer conflicting opinions the court is not to decide which expert is correct”). Contrary to one critic’s metaphor, the judge’s task is *not* like having to decide which of “two groups of expert mathematicians [who] disagree about a complex mathematical question” is correct. Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 *Yale L.J.* 1535, 1538 (1998). In tort and criminal cases, the judge’s task is not to decide whether a given solution or theory is correct, but to decide whether, given the current state of knowledge about the solution or theory as propounded by competing experts, it is sufficiently trustworthy to come into evidence in order to prove or disprove a legal issue. Examples of legal issues which scientific hypotheses could help prove or disprove include whether a defendant pharmaceutical company’s drug could have caused the plaintiff’s injuries, or whether a criminal defendant’s presence could be established at the scene of the crime.

of science.¹⁹ While judges do need a mental framework for this analysis, they do not need to be trained scientists.

Evidentiary decisionmaking thus falls squarely within the increasingly important realm of interdisciplinary studies.²⁰ This Article seeks to expand the interdisciplinary tradition by developing a conceptual framework enabling legal actors to approach questions of scientific evidence in a logical, analytic fashion.²¹ It seeks to illustrate unifying themes common to scientific thinking of all stripes. I assert that understanding the language and structure of scientific argument and the way “science” is produced provides an invaluable tool in deciphering the logic behind the testimony.

The framework proposed in this Article is intended to resolve some major issues on which the courts are still foundering. In so doing, it seeks to equip judges with the tools necessary to provide “intellectual due process”—a structured reasoning process that is not arbitrary—to the parties in cases involving scientific evidence.²² This

¹⁹ See *United States v. Hall*, 165 F.3d 1095, 1102 (7th Cir. 1999) (“[W]hen faced with a proffer of expert scientific testimony, the district court must ‘consider whether the testimony has been subjected to the scientific method; it must rule out subjective belief or unsupported speculation.’” (quoting *Porter v. Whitehall Labs., Inc.*, 9 F.3d 607, 614 (7th Cir. 1993) (citations omitted))). See generally Arthur Fine, *The Shaky Game: Einstein, Realism, and the Quantum Theory* (1986) (asserting that questions about truth claims of science must be answered by reference to methods and standards of science).

²⁰ As George Priest noted, “the increase of interdisciplinary ‘law and’ scholarship . . . refine[s] our understanding of the law.” George L. Priest, *The Growth of Interdisciplinary Research and the Industrial Structure of the Production of Legal Ideas: A Reply to Judge Edwards*, 91 Mich. L. Rev. 1929, 1934 (1993); see also Robert J. Condlin, “What’s Really Going On?” A Study of Lawyer and Scientist Inter-Disciplinary Discourse, 25 Rutgers Computer & Tech. L.J. 181, 182-83 (1999) (“[S]cience and law have become pervasively intertwined . . . Science based choices and problems pervade the most difficult issues of social regulation facing legislatures and courts.”).

²¹ To be valuable to society, lawyers and judges must have a grasp of the issues presented in litigation, a growing number of which involve questions of scientific evidence. See David L. Faigman, *Legal Alchemy: The Use and Misuse of Science in the Law* at xiii (1999) (“[S]ince law must rely on science, it is incumbent on lawyers and policy makers to understand it.”); Harry Kalven, Jr., *The Quest for the Middle Range: Empirical Inquiry and Legal Policy*, in *Law in a Changing America* 56, 58 (Geoffrey C. Hazard, Jr. ed., 1968) (arguing that “we in the legal world need some literacy as to scientific method”).

²² See Brewer, *supra* note 18, at 1539 (arguing that intellectual due process is norm that “places important *epistemic* constraints on the reasoning process by which legal decisionmakers apply laws to individual litigants . . . [and requires] that the decisionmaking process not be arbitrary from an epistemic point of view”). Professor Scott Brewer developed the concept of intellectual due process and argues that intellectual due process is incompatible with the way most judges currently decide questions of scientific validity. See *id.* at 1539. I agree with Professor Brewer’s criticism that judges’ “reliance on such indicia of expertise as credentials, reputation, and demeanor” instead of scientific understanding “yield[s] only *epistemically arbitrary* judgments.” *Id.* I do not, however, share Brewer’s pessimism with regard to the abilities of the judiciary to learn how to assess scientific argument. See *id.* at 1680 (concluding that “unless judges are routinely and *systematically* trained in scientific theories and methods, *Daubert* does not offer a promising overall solu-

Article seeks to empower judges to utilize the approach required by *Daubert* to provide the intellectual due process necessary in every case.²³

The heuristic proposed here consists of five basic parts and emphasizes the underlying principles common to all fields of science. To meet the requirements of intellectual due process in making admissibility determinations, I suggest that judges (and the lawyers who educate them about their cases) must be able to do five things: (1) identify and examine the proffered theory and hypothesis for their power to explain the data; (2) examine the data that supports (and undermines) the expert's theory; (3) use supportable assumptions to fill the inevitable gaps between data and theory;²⁴ (4) examine the methodology; and (5) engage in probabilistic assessment of the link between the data and the hypothesis.²⁵

tion"). Indeed, Brewer acknowledges that "[n]othing in [his] Article suggests that a nonexpert judge cannot become sufficiently epistemically competent, even without the formal training of a scientist." *Id.* This Article provides a framework to enable judges to achieve just such epistemic competence.

²³ Throughout this Article, I argue that judges are capable of providing intellectual due process to litigants on issues of scientific evidence. It is imperative, however, that judges educate themselves about the kinds of evidence before them, make default assumptions that are justifiable on scientific and policy grounds, and explicitly give the basis for their decision in the form of a written opinion.

²⁴ The reliance on assumptions within the heuristic may strike many legal readers as a departure from legal concepts of causation. In science, however, background theories and assumptions underpin any theory, hypothesis, or experiment. See 1 Schum, *supra* note 16, at 35 (noting necessity and importance of assumptions in scientific argument). Being able to identify and critique these background assumptions is key to understanding scientific argument. See *infra* notes 85-92 and accompanying text for discussion of the role of background assumptions.

²⁵ This Article asserts that the underlying principles of reasoning in law and science are not different, although context and culture determine their application. I include inductive, deductive, and abductive reasoning in this statement. All three forms of reasoning are important tools in analysis. See 1 Schum, *supra* note 16, at 18-20 (offering elegant explanation of inductive, deductive, and abductive reasoning and process of logical inference). In short, for deductive argument to be valid, the truth of the premises must guarantee the truth of the conclusion. See *id.* at 18. The paradigmatic form of the deductive argument is the Aristotelian syllogism. See Marjorie Grene, *A Portrait of Aristotle* 67 (1963). By inductive reasoning, I mean both inductive generalization, involving probabilistic generalization from the particular, and inductive analogy, in which one concludes that some particular instance will have the aggregate characteristics given in the premises. See Stephen F. Barker, *The Elements of Logic* 223-31 (2d ed. 1974). Exemplary reasoning is sometimes referred to as abduction. For an article describing the process of legal reasoning by analogy as a species of abduction, see Scott Brewer, *Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument by Analogy*, 109 *Harv. L. Rev.* 923, 947 (1996), arguing that

[A]bduction is a disciplined (albeit, in contrast to deduction, not a rigidly guided) form of inference; . . . it has a substantial degree of rational force; and . . . it plays a vital role in exemplary, analogical reasoning, just as it does in explanatory and justificatory reasoning in science and other fields of inquiry.

It is my hope in designing this heuristic that judges will be able to make better admissibility decisions. The common law's long-standing requirement that judges explain the basis of their reasoning and that the basis be well-founded forms the cornerstone of a system that aspires to rationality.²⁶ This Article seeks to equip judges—and the lawyers who must practice before them—in that endeavor.

Part I of this Article presents an overview of scientific argument, discussing the meaning scientists attach to scientific method and its rhetorical use. Building on this understanding of scientific argument and analysis, Part II presents the proposed heuristic, and explains each of its steps in detail. Part III applies the heuristic in two cases, a toxic tort case and a criminal case.

I

EVALUATING SCIENTIFIC ARGUMENT: WHAT IS THE SCIENTIFIC METHOD ANYWAY?

According to the Supreme Court, whether testimony amounts to admissible scientific knowledge depends on whether it has been “derived by the scientific method.”²⁷ The Supreme Court provided four guidelines for this inquiry: testability, peer review and publication, error rate and the existence of protocols, and general acceptance.²⁸ Other than setting out its four “flexible” guidelines, however, the Supreme Court gave federal judges no inkling as to what the “scientific method” was all about.²⁹ Although most of us have heard the term “scientific method,” and *Webster's Dictionary* defines it,³⁰ expecting judges to decide whether expert testimony has been derived from its criteria—without more guidance—seems a stretch. Federal judges now know that it is their job to determine the scientific validity of

Charles Sanders Peirce introduced the theory of abduction to explain how scientists select a relatively small number of hypotheses to test from a large number of logically possible explanations for their observations. See Charles S. Peirce, *Philosophical Writings of Peirce* 150-56 (Justus Buchler ed., 1955).

²⁶ See Brewer, *supra* note 25, at 930 (“The normative order constituted by the legal system, informed by ‘rule of law’ principles as well as by many others, aspires to be rational in significant ways.”).

²⁷ *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590 (1993).

²⁸ See *id.* at 593-94. For a critique of these guidelines, see Beecher-Monas, *supra* note 8, at 1052-55.

²⁹ See *Daubert*, 509 U.S. at 590, 594.

³⁰ *Webster's Dictionary* defines “scientific method” as “principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses.” *Webster's Ninth New Collegiate Dictionary* 1051 (1989).

expert testimony,³¹ and that to perform this task, they must refer to the scientific method.³² But what is the scientific method, anyway?

A. *Scientific Method: The Myth*

As commonly understood, “scientific method” means taking an idea about how things work, framing it as a testable hypothesis, and testing the hypothesis to see if it holds true, all measured and expressed in mathematical—probabilistic—terms.³³ Despite the Supreme Court’s reference to the scientific method as the cornerstone of scientific knowledge, however, there is no monolithic “scientific method.”³⁴ Karl Popper, the preeminent philosopher of science cited by the *Daubert* Court,³⁵ thoroughly debunked any such notion.³⁶ There is no all-inclusive set of rules that can be applied to science to determine its validity.³⁷ Differences arise among scientists in different

³¹ See supra note 3 and accompanying text (discussing *Daubert*).

³² The content of the scientific expert’s testimony must be “scientific knowledge.” *Daubert*, 509 U.S. at 590. To be scientific knowledge, the testimony must be “derived by the scientific method.” *Id.* In *Kumho Tire*, the Supreme Court extended the requirements of *Daubert* to all expert testimony, including nonlaboratory sciences such as engineering. See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 147 (1999). Understanding scientific method as a form of argument applicable to all forms of expert testimony makes this Article relevant even to technical testimony, and highlights the flimsiness of the whole dichotomy between “technical” and “scientific” testimony. See Beecher-Monas & Garcia-Rill, supra note 5, at 249 (“Understanding the methodology, requiring the existence and maintenance of standards controlling the technique’s operation, and examining how often a given procedure yields mistaken results, are crucial to evaluating an expert’s conclusions.”). Testimony of all kinds either has empirical support or it does not. The type of data may be different, as may be the means of gathering it, but the argument still must meet the requirements of intellectual honesty, which is the essence of the scientific method. See supra note 5 and accompanying text (discussing *Kumho Tire*).

³³ See Charles Van Doren, *A History of Knowledge 188-89* (1991) (describing this version of scientific method, termed “experiment,” as “[t]he best known method, but not necessarily the most often employed”).

³⁴ See Beecher-Monas & Garcia-Rill, supra note 5, at 254 (noting that for Popper, “there was no such thing as the ‘scientific method’”).

³⁵ See *Daubert*, 509 U.S. at 593 (citing Popper for belief that testability is key to determining theory’s standing).

³⁶ See Karl R. Popper, *The Logic of Scientific Discovery* 278 (rev. ed. 1992) (explaining that science is not system of certainty, or well-established statements, nor can it ever attain truth, or even probability).

³⁷ Occasionally, commentators argue that the postulates proposed by Robert Koch and Friedrich Gustav Jacob Henle, or those by Austin Bradford Hill, can be used as criteria for good science. See, e.g., Kenneth R. Foster & Peter W. Huber, *Judging Science* 28 (1997). Hill’s postulates require consideration of strength of association, consistency with other scientists’ results, specificity of association, temporality, biological gradient or dose-response curve, biological plausibility, coherence, and analogy with similar evidence. See *id.* at 30-31. Koch and Henle outlined 10 criteria for causation, including (1) higher prevalence of disease in exposed than unexposed populations, (2) those with the disease should have had more exposure to the agent than healthy populations, (3) experiments should demonstrate increased incidence of disease in exposed as opposed to unexposed popula-

disciplines, and even within the same discipline, with respect to the amount of evidence needed to support a theory.³⁸ Instead of a monolithic scientific method, there are many methods and many procedures.³⁹ The only essential ingredient for good science—and hence

tions, (4) temporality, (5) linear dose-response curve, (6) biological plausibility, and a number of other factors relating to immune responses. See Alfred S. Evans, *Causation and Disease: The Henle-Koch Postulates Revisited*, 49 *Yale J. Biology & Med.* 175, 192 tbl.13 (1976); see also Foster & Huber, *supra*, at 28 (describing Evans's list as "modern version" of Koch's postulates). But while these respected canons are unquestionably useful in designing optimal experiments, they cannot be—and were not intended to be—used as a checklist for sound science. See Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 *Proceedings Royal Soc'y Med.* 295, 295-300 (1965), reprinted in *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods* 15-20 (Sander Greenland ed., 1987); see also Foster & Huber, *supra*, at 28 (noting that "[m]ost scientists would agree that evidence satisfying all Koch's postulates establishes a compelling case" for causation, but explaining that they are neither necessary nor sufficient conditions for inferring causation); Evans, *supra*, at 192 (explaining that Koch's postulates "were not regarded as rigid criteria by Koch himself and should not be today").

³⁸ For an example, consider the debate among scientists over which chemicals are human carcinogens. See Bruce N. Ames & Lois Swirsky Gold, *Too Many Rodent Carcinogens: Mitogenesis Increases Mutagenesis*, 249 *Science* 970, 970-71 (1990) (questioning animal studies-based evidence showing chemicals to be carcinogens in humans). But see Jean Marx, *Animal Carcinogen Testing Challenged*, 250 *Science* 743 (1990) (explaining theories of Ames's critics that animal studies conducted to date provide sufficient evidence of carcinogenicity of chemicals in humans).

³⁹ Unfortunately for judges searching for clear rules to distinguish good science from bad, there simply are none. Even good studies may not meet all the possible criteria for a "perfect" study. Hill cautioned: "I do not believe . . . that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect." See Hill, reprinted in *Evolution of Epidemiologic Ideas*, *supra* note 37, at 19 (discussing his evidentiary factors used to prove that particular agent caused disease). In fact, none of the factors Hill proposed provided "indisputable evidence for or against the cause-and-effect hypothesis." *Id.* For example, though most scientists would place great weight on Hill's criterion of consistency of observed effect, it may be absent even from strong causal connections. See Sander Greenland, *Preface to Hill*, reprinted in *Evolution of Epidemiologic Ideas*, *supra* note 37, at 14, 14 (noting that "a careful reading of Hill shows that he did not intend to offer a list of *necessary* conditions" and that of Hill's factors, only temporality is necessary (emphasis added)).

While finding a consistent specific association between exposure and a particular disease might be persuasive, such an association is rarely observed. See Hill, *supra* note 37, at 297 (noting that finding "specificity" greatly strengthens conclusion of causation but is not required in order to make conclusions). These strong specific associations are often referred to by courts as a "signature" disease, although most courts recognize that such associations are not a prerequisite for admissibility. See, e.g., *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1402 (D. Or. 1996) (examining causation evidence in silicone exposure case despite absence of "signature disorder" for silicone exposure, although ultimately excluding evidence); *Maiorana v. National Gypsum Co. (In re Joint Dist. Asbestos Litig.)*, 827 F. Supp. 1014, 1046 (S.D.N.Y. 1993), *rev'd on other grounds*, 52 F.3d 1124 (2d Cir. 1995) (acknowledging that mesothelioma is "signature disease" for asbestos, but still examining causal relationship between plaintiffs' *colon cancer* and asbestos exposure). A linear dose-response curve, indicating a positive relationship between the administered dose and its effect (i.e., the higher the dose, the stronger the effect), is similarly important, but the absence of such a response does not destroy necessarily the causal inference, al-

the only overarching method—is that good science must be open to critique and revision.⁴⁰ Scientists recognize that what matters most is the explanatory power of the proffered theory and how well the data support the theory.⁴¹

Purportedly, the scientific method is a way to ensure objectivity.⁴² But the scientific method is no guarantee of objectivity.⁴³ Although empirical content is fundamental to science, the view that facts are “immanent in an objective reality waiting to be discovered by any scientists who look in the right place” is just as false in science as it is in law.⁴⁴ Mere observation of the world does not lead to scientific

though it does indicate a more complex relationship between the agent and disease. See Hill, reprinted in *Evolution of Epidemiologic Ideas*, supra note 37, at 18.

The Federal Judicial Center’s Reference Manual accepts this proposition, noting that although the presence of a linear dose-response relationship “strengthens the conclusion that the relationship . . . is causal[,] . . . a dose-response relationship is not necessary to infer causation.” Linda A. Bailey et al., *Reference Guide on Epidemiology*, in *Federal Judicial Ctr. Manual*, supra note 9, at 121, 164. The one exception to the notion that the Hill criteria need not be met for causation to exist is that the temporality factor—the notion that cause should precede effect—is universally recognized as a prerequisite for admissibility. Yet all agree that temporality alone is not enough. See Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 16 Va. Envtl. L.J. 1, 43 (1996) (stating that temporality “has become recognized as the one factor which should actually be viewed as a requirement for the admissibility of studies”); see also Greenland, supra, at 14 (noting that temporality is only Hill factor considered necessary to establish causation). As an example of the insufficiency of temporality alone, consider that just because most dead men are bald does not necessarily mean that baldness is fatal.

⁴⁰ See Popper, supra note 36, at 279 (stating that scientists’ conjectures, once proffered, are not “dogmatically upheld,” but rather scientists attempt to “overthrow them”). This ability of good science to withstand constant critique is what Popper called “falsifiability,” a concept that the Supreme Court emphasized in *Daubert*. See *Daubert*, 509 U.S. at 593 (citing Popper and noting that falsifiability is key in determining whether proffered evidence constitutes scientific knowledge).

⁴¹ See Popper, supra note 36, at 108, 111 (stating that best theory “holds its own in competition with other theories” and that there is “nothing ‘absolute’ about [science]”).

⁴² See id. at 44 (explaining Kant’s idea that scientific justification consists of objective verification (citing Immanuel Kant, *Critique of Pure Reason* 645 (N. Jemp Smith transl., 1933))); cf. Sean O’Connor, *The Supreme Court’s Philosophy of Science: Will the Real Karl Popper Please Stand Up?*, 35 *Jurimetrics* 263, 266-67 (1995) (explaining Kantian objectivist view that we “impose a logical order and structure onto our experience of the world”).

⁴³ See Karl R. Popper, *The Myth of the Framework* 8 (M.A. Notturmo ed., 1994) (“There is no pure, disinterested, theory-free observation.”).

⁴⁴ Peter H. Schuck, *Multi-Culturalism Redux: Science, Law and Politics*, 11 *Yale L. & Pol’y Rev.* 1, 15 (1993).

knowledge.⁴⁵ Observation must be informed by theory.⁴⁶ Science is a creative enterprise.⁴⁷

Thus, science is about more than a mere accumulation of facts. But, to make matters more difficult for judges seeking a set of rules for assessing science, just as “facts” are intertwined with subjective values in legal decisionmaking,⁴⁸ scientific conclusions are also based on subjective judgments made at key points ranging from the initial decision to study a particular phenomenon through the collection, categorization, and interpretation of data.⁴⁹ Interpreting experiments is neither a simple nor a disinterested process.⁵⁰ On the contrary, scientific paradigms are—like legal paradigms—socially constructed⁵¹

⁴⁵ See Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* 161 (1958) (“Factuality is not science. Only a comparatively few peculiar facts are scientific facts, while the enormous rest are without scientific interest.”).

⁴⁶ See Popper, *supra* note 36, at 107 (“Theory dominates the experimental work from its initial planning up to the finishing touches in the laboratory.”). A striking example is given by a leading scientist: “[E]ven the so-called ‘fact’ that a star is located in a certain position in the heavens is a consequence of Einstein’s *theory* which determines how much light may deviate from a straight line.” Stephen G. Brush, *Prediction and Theory Evaluation: The Case of Light Bending*, 246 *Science* 1124, 1126 (1989).

⁴⁷ Although we think of scientific facts as having been discovered, rather than fabricated, the etymology of the word “fact” is from the Latin *facere*, which means “to make.” See Karin D. Knorr-Cetina, *The Manufacture of Knowledge* 3 (1981). As one scientist explained, “[m]ost of the reality with which scientists deal is highly preconstructed, if not wholly artificial.” *Id.* For an excellent discussion of how theory and fact intertwine, see generally Harold I. Brown, *Perception, Theory and Commitment* (1979).

⁴⁸ See, e.g., William Twining, *Rethinking Evidence* 107 (1990) (noting that since “triers of fact are regularly and unavoidably involved in making evaluations . . . it is misleading to suggest that legal enquiries into questions of fact are value-free”).

⁴⁹ See Thomas S. Kuhn, *The Structure of Scientific Revolutions* 200 (2d ed. 1970) (arguing that manner in which science develops is largely determined by values and experience shared by scientific community); see also David L. Faigman, *To Have and Have Not: Assessing the Value of Social Science to the Law as Science and Policy*, 38 *Emory L.J.* 1005, 1028 (1989) (“Scientists select problems on the basis of what seems important, and to this extent all science is culture-bound.”); Michael L. Seigel, *A Pragmatic Critique of Modern Evidence Scholarship*, 88 *Nw. U. L. Rev.* 995, 1035 (1994) (“[T]he scientific method necessarily encompasses a multitude of value-laden judgments that affect the outcome of research.”).

⁵⁰ Although the systematic testing of proposed explanations of reality is a basic feature of scientific process, it is the theoretical construct—the hypothesis—which precedes and pervades experimentation. See Popper, *supra* note 36, at 107 n.*3. An example of this is the emptiness of the data dredging experiments, in which computers were employed to elicit huge quantities of “facts” that no one yet has been able to understand. Cf. Patricia K. Woolf, *Deception in Scientific Research*, in *American Ass’n for the Advancement of Science, Project on Scientific Fraud and Misconduct: Report on Workshop Number One* 37, 78 (1988) (describing data dredging as “making inappropriate or unwarranted inferences based on unreported statistical techniques”); Edward L. Rubin, *The Practice and Discourse of Legal Scholarship*, 86 *Mich. L. Rev.* 1835, 1899 n.163 (1988) (“Without a methodology, fact gathering becomes a mechanical or incoherent process that has little intellectual appeal.”).

⁵¹ As an initial matter, the availability of funding for particular experiments—clearly a social phenomenon reflecting exigencies of politics and demand—determines which theo-

through a process of discussion and consensus-building about theories, experimental methods, instrumentation, and validation.⁵² This does not make them any less reliable, however.⁵³

Acknowledging that science is a creative process does not imply that everything is up for grabs.⁵⁴ There are shared perceptions of validity. That is, even if there is no set of inflexible "rules" for judging the validity of a scientific argument, there are criteria for formulating and testing scientific ideas.⁵⁵ To be defined legitimately as science, a hypothesis or argument still must have persisted despite radical

ries will be examined and which data analyzed. In addition, what we come to know is inextricably a social process. See Michel Foucault, *Power/Knowledge* 69 (Colin Gordon ed., Colin Gordon et al. trans., 1980) ("Once knowledge can be analysed in terms of region, domain, implantation, displacement, transposition, one is able to capture the process by which knowledge functions as a form of power and disseminates the effects of power."). Further, what will count as "truth" is a matter of negotiation between various factions that have a stake in the matter. Thus, the definition of "scientific" fact "must be sought through interpretation of the changing borders and territories of what is taken as science." Thomas F. Gieryn, *Boundaries of Science*, in *Handbook of Science and Technology Studies* 393, 417 (Sheila Jasanoff et al. eds., 1995). For example, the feminist critique of science studies the shifts in the boundary of what counts as science toward the masculine and away from the feminine, claiming that not only are women underrepresented in science, but scientific research underrepresents women. See *id.* at 420-24.

⁵² See Sheila Jasanoff, *What Judges Should Know About the Sociology of Science*, 32 *Jurimetrics* 345, 347 (1992) (explaining view that "the 'facts' that scientists present to the rest of the world are not simple reflections of nature; rather, these 'facts' are produced by human agency, through the institutions and processes of science, and hence they invariably contain a social component"). Thomas Kuhn's philosophy of science provides useful insight into the sociology of scientific paradigm change. See generally Kuhn, *supra* note 49. For a view that, while science does reveal "reality," it does so by the process of consensus and community activity, see generally Henry H. Bauer, *Scientific Literacy and the Myth of the Scientific Method* (1992). It does not follow from the consensual nature of the enterprise that scientific conclusions are completely indeterminate, however. See J.M. Ziman, *Public Knowledge* 9 (1968) (arguing that nature constrains science's goal of reaching consensus of rational opinion). The shared perceptions regarding how to determine experimental and technical validity provide some constraints, just as shared perceptions of accepted judicial methodology provide restraints on judicial activism.

⁵³ See Popper, *supra* note 43, at 13 (observing that "the objectivity and the rationality of progress in science [are] not due to the personal objectivity and rationality of the scientist," but to their ability to stand up to critical examination).

⁵⁴ See *id.* at 33 (criticizing notion that truth is relative to intellectual background).

⁵⁵ Cf. Popper, *supra* note 36, at 313-14 (explaining that theories only convey information to extent they are capable of being systematically tested). These shared perceptions of validity depend on testing the internal consistency of the theory, the consistency of the theory with other theories, and the consistency of the theory with the empirical data. See *id.* at 315. Most importantly, validity requires that science continually question and criticize all of its theories, constantly refining its explanations. See *id.* at 317 (stating that [W]e must not look upon science as a 'body of knowledge,' but rather as a system of hypotheses . . . with which we work as long as they stand up to tests, and of which we are never justified in saying that we know that they are 'true' or 'more or less certain' or even 'probable.'").

For an explanation of Popper's philosophy, see Beecher-Monas & Garcia-Rill, *supra* note 5, at 255 ("[T]he aim of science is to achieve better and better explanations.").

changes in its content.⁵⁶ Such science consists of background assumptions about the way the world works, coupled with canons of experimental design and theoretical exemplars or models that translate background assumptions into working rules that guide the selection of problems and inform acceptable procedures for their solution.⁵⁷ Explicitly acknowledging the subjective element inherent in all scientific methodologies thus does not make science an arbitrary enterprise, but instead opens for questioning the underlying assumptions of a particular scientific argument. To determine what is really at stake when experts disagree, it is necessary to examine those underlying assumptions.⁵⁸

B. *Scientific Method: The Reality*

Whether or not the scientific method exists in practice, scientists and nonscientists alike frame their arguments about scientific issues as though it does.⁵⁹ Even Popper did not dispute that there are indeed good methods in science.⁶⁰ He merely cautioned against sanctioning any rules that would exclude alternate forms of inquiry.⁶¹ Thus, scientists continue to describe their evaluation process generally in terms of scientific method, despite their skepticism.⁶²

⁵⁶ See Kuhn, *supra* note 49, at 2-3 (explaining that since “out-of-date theories are not in principle unscientific because they have been discarded,” science includes “bodies of belief quite incompatible with the ones we hold today”).

⁵⁷ See *id.* at 200.

⁵⁸ Every hypothesis may rely on an underlying set of assumptions, or “auxiliary hypotheses,” that in turn must be justifiable. See, e.g., O’Connor, *supra* note 42, at 274 (discussing “auxiliary hypotheses,” and critiquing Popper for his perceived failure to address this problem). Popper does address the problem of using assumptions. See Popper, *supra* note 36, at 83, 104-05. However, his critics (including O’Connor and Lakatos) contend—wrongly, in this author’s opinion—that Popper largely ignored the implications. See Irme Lakatos, *The Methodology of Scientific Research Programmes* 49 (1978) (explaining importance of requiring assumptions to have empirical content); *id.* at 80 n.4 (discussing use of assumptions). On the contrary, Popper acknowledged the inevitability of using assumptions, but required that assumptions also have falsifiable consequences. See Popper, *supra* note 36, at 83 (explaining that assumptions or “auxiliary hypotheses” may be introduced to assist theory, but also must add explanatory value).

⁵⁹ Although the term “scientific method” has been criticized by philosophers of science as too monolithic to have meaning, scientific argument still proceeds from this basic concept while taking issue with it. See Alan G. Gross, *The Rhetoric of Science* 32 (1990).

⁶⁰ See Popper, *supra* note 36, at 42 (“[W]hat characterizes the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested.”).

⁶¹ See *id.* at 54 (explaining importance of designing scientific rules of procedure “in such a way that they do not protect any statement in science against falsification”); Beecher-Monas & Garcia-Rill, *supra* note 5, at 254 (describing how Popper’s reluctance to embrace standard scientific method stemmed from belief that alternate forms of inquiry are beneficial).

⁶² See, e.g., Stephen David Ross, *The Scientific Process* 81 (1971) (“Science, as a mode of human activity, is dominated by a conception of what it means to explain events through prediction and control, by collecting evidence for proposed hypotheses, by solving certain

Moreover, even without a single, universal scientific method, there is consensus about certain basic principles.⁶³ Probabilistic assessment of data, independent verification of new procedures, and concern over false positives are fundamental to scientific argument in every discipline.⁶⁴ Metaphor is also basic to scientific understanding in every field.⁶⁵ In addition, quantum physics has added the insight that precise prediction of future behavior is impossible; the most that can be predicted is the probability of various behaviors.⁶⁶ Principles of complexity theory—which state that complex systems consist of interacting parts capable of shaping themselves into organized fluctuating patterns—have replaced deterministic notions of causation.⁶⁷ These ideas form the background of much modern scientific discourse. Thus, the reality of scientific method is a common understanding of the centrality of probabilistic reasoning; the importance of testability, interdisciplinarity, and rationality; and an emphasis on the explanatory power of a proposed hypothesis.

kinds of problems.”); cf. Gross, *supra* note 59, at 32 (arguing that although “there is no scientific method, no global strategy for all of science[,] . . . there are scientific methods, the aggregate of agreed-upon procedures”).

⁶³ Although most procedures have limits, and therefore detractors, and no method is universally accepted, argument nonetheless proceeds from certain basic understandings about these procedures. See Gross, *supra* note 59, at 32 (explaining that “commitment in science . . . rests ultimately on agreed-upon procedures”).

⁶⁴ See, e.g., Lincoln E. Moses, *Statistical Concepts Fundamental to Investigations, in Medical Uses of Statistics* 5, 14-26 (John C. Bailar III & Frederick Mosteller eds., 2d ed. 1992) (applying probabilistic thinking to medical analyses). As noted earlier, probabilistic assessment of data means interpreting results of a study in terms of relative frequencies in repeated trials under controlled conditions. See *supra* note 14 (discussing long-run relative frequency). Independent verification of new procedures requires that a methodology be repeatable by another scientist to achieve similar results. A false positive means that a result is claimed or observed that actually did not occur. Scientists would prefer to err on the side of claiming no result when there actually was one rather than the converse.

⁶⁵ See generally Gross, *supra* note 59, at 47 (discussing centrality of metaphor to science and explaining that scientific models should be understood as metaphors); see also Michael R. Matthews, *Science Teaching: The Role of History and Philosophy of Science* 205 (1994) (observing that “[a] good deal of science method courses are devoted to practice in the use of metaphor and models”). Indeed, Lakoff & Johnson argue that metaphor is indispensable to reasoning in any human endeavor. See George Lakoff & Mark Johnson, *Metaphors We Live By* 221 (1980) (“The use of many metaphors that are inconsistent with one another seems necessary for us if we are to comprehend the details of our daily existence.”).

⁶⁶ For example, the exact position of electrons cannot be determined. See Beecher-Monas & Garcia-Rill, *supra* note 5, at 253 n.51 (contrasting predictability of trajectories of Newtonian mechanics with unpredictability of wave functions in quantum physics, where exact location of electron even in simple system is impossible to describe).

⁶⁷ See *id.* at 252-53 (comparing Newtonian determinism, which suggested that, once initial conditions for event were known, it was possible to calculate not only what would happen next, but what had happened before, with modern understandings of fluctuations, instability, self-organization, multiple choices, and limited predictability). For an overview of complexity theory, see generally M. Mitchell Waldrop, *Complexity* (1992).

1. Probabilistic Reasoning

Probabilistic reasoning underlies modern science, especially science involving complex systems like biological systems.⁶⁸ By probabilistic reasoning, I mean a way of thinking about the physical characteristics of an object or event.⁶⁹ I do not mean subjective probabilities⁷⁰ as prediction about whether a theory is true or not.⁷¹ Although scientists may attach subjective probabilities to their conclusions, for modern scientists, probability is a quantitative measure of chance, a physical quality just like weight or shape, and has nothing to do with the personal beliefs of the scientist.⁷² Scientists understand

⁶⁸ See Ian Hacking, *Was There a Probabilistic Revolution 1800-1930?*, in 1 *The Probabilistic Revolution* 45, 45 (Lorenz Kruger et al. eds., 1987) (arguing that "today our vision of the world is permeated by probability"); see also Troyen A. Brennan, *Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation*, 73 *Cornell L. Rev.* 469, 483 n.70 (1988) (observing that "testing and rejection of hypotheses based on statistical analysis is integral to science and offers a more realistic paradigm of scientific thought than does pure deduction").

⁶⁹ See Jan von Plato, *Probabilistic Physics the Classical Way*, in 2 *The Probabilistic Revolution*, supra note 68, at 379, 379-80 (explaining statistical probability as physical property of macroscopic systems rather than indication of likelihood or degree of belief).

⁷⁰ Theodore Colton proposed the probability of the President of the United States dying in office as a useful example to illustrate the difference between subjective and frequentist probability. This probability, Colton suggested, either can reflect the consideration that the President is one of a group of males of a particular age living in a particular geographic locale (the frequentist approach), or it can mean one's *belief* in the likelihood of the President's death (the subjectivist approach). See Colton, supra note 14, at 63. This Article adopts the frequentist approach.

⁷¹ Probabilistic reasoning does not mean that one can say a theory is more or less probable. See Karl R. Popper, *Objective Knowledge* 55 (1981). Rather, it refers to an understanding that probability (or propensity, as Popper called it) is as much a property of the object being observed as is its mass and exists independently of anyone's belief. See Ian Hacking, *The Emergence of Probability* 14 (1975) ("The propensity to give heads [on a coin flip] is as much a property of the coin as its mass, and the stable long run frequency found on repeated trials is an objective fact of nature independent of anyone's knowledge of it, or evidence for it.").

⁷² See Colton, supra note 14, at 71 (noting that Bayesian ideas have not supplanted frequentist methods of statistical inference). Under a subjectivist (or Bayesian) paradigm, probability denotes a degree of expectation (or subjective certainty). See Andreas Kamlah, *The Decline of the Laplacian Theory of Probability: A Study of Stumpf, von Kries, and Meinong*, in 1 *The Probabilistic Revolution*, supra note 68, at 91, 112; see also Colin Howson & Peter Urbach, *Scientific Reasoning: The Bayesian Approach* 11 (1989) (acknowledging, although their viewpoint is Bayesian, that frequentist "classical methods of statistical inference [] have achieved pre-eminence" in science). But, as Popper points out, "it is absurd to explain objective statistical frequencies by subjective ignorance." Karl R. Popper, *Quantum Theory and the Schism in Physics* 106 (W.W. Bartley, III ed., Rowan & Littlefield 1982) (1956); see also Richard Bevan Braithwaite, *Scientific Explanation* 120 (1953) (explaining that "the probability of disintegration of a radium atom contains no reference to . . . belief"). But see Pagano & Gauvreau, supra note 14, at 135-40 (suggesting that Bayes Theorem may be useful for measuring accuracy of medical screening and diagnosis).

that “fluctuations, instability, multiple choices, and limited predictability” are inherent “at all levels of observation.”⁷³

In other words, it is inconsistent with scientific notions of probability to say that a theory is more likely true than not.⁷⁴ Scientists understand that there is simply no way to prove a theory true or even probable, as any theory may be subject to being disproved.⁷⁵ So the most that can be said is that a particular theory has withstood criticism and provides the best explanation of the data.⁷⁶

Probabilistic reasoning encompasses the idea that statistical laws are fundamental to scientific explanation.⁷⁷ Statistical inference, with its concepts of probability, independence, and randomness, is basic to scientific discourse and forms the cornerstone of data analysis as well as the basis for causation and other explanatory arguments.⁷⁸ Funda-

⁷³ Beecher-Monas & Garcia-Rill, *supra* note 5, at 253.

⁷⁴ See Braithwaite, *supra* note 72, at 123 (noting that scientific hypothesis cannot be established as true because “however many observations we have made[,] a further observation may serve to refute the hypothesis”); see also Popper, *supra* note 36, at 316-17 (discussing distinction between probability theory and probability of hypothesis). Many scientists vehemently believe in their theories. My point is that they understand that there may be a disconnect between their belief and the reality of science in practice. Cf. Hacking, *supra* note 71, at 14 (recognizing that majority of practitioners ignore distinction between objective and subjective probability).

⁷⁵ See Stephen F. Lanes, *Error and Uncertainty in Causal Inference*, in *Causal Inference* 173 (Kenneth J. Rothman ed., 1988). As Stephen Lanes wrote, “there can be no empirical support for the conclusion that causation is the ‘most likely’ explanation for any association.” *Id.* at 183. He added that “[t]he uncertainty in causal inference is attributable to the fact that we cannot establish that an association is valid.” *Id.* at 185. An unidentifiable error may exist and it may cause the observation. See *id.* The most that can be expected of probability factors such as “strength of an association, the shape of a dose-response curve, [and] the level of statistical significance” is that they “affect subjective beliefs.” *Id.* at 186. Truth, in other words, must be distinguished from its fallible signs. See Irme Lakatos, *Mathematics, Science and Epistemology* 108 n.2 (1978) (explaining that “a proposition may correspond to a fact but there is no infallible way to establish this correspondence”).

⁷⁶ See Beecher-Monas & Garcia-Rill, *supra* note 5, at 254-55. As Patrick Suppes explains, “probability theory is designed to discover and to analyze partial causes in complex situations for which a complete causal analysis is not feasible.” Patrick Suppes, *A Probabilistic Theory of Causality* 8 (1970).

⁷⁷ See Lorenz Kruger, *The Slow Rise of Probabilism: Philosophical Argument in the Nineteenth Century*, in 1 *The Probabilistic Revolution*, *supra* note 68, at 59.

⁷⁸ See Michael O. Finkelstein & Bruce Levin, *Statistics for Lawyers* 107 (1990) (“Statistical inference rests on concepts of probability and randomness.”). As Suppes explains, “[t]he concept of independence is one of the most profound and fundamental ideas, not only of probability theory but of science in general.” Suppes, *supra* note 76, at 109. The concept of independence explains that discrete biological or physical events, like coin tosses, have no memory of time and place. See Lynn Arthur Steen, *The Science of Patterns*, 240 *Science* 611, 615 (1988). In other words, knowing whether a coin toss yielded heads or tails on the previous flip will not help predict what the next coin toss will yield. See Frederick Mosteller et al., *Probability with Statistical Applications* 31 (2d ed. 1970). On the other hand, an idealized “fair” coin has a probability of yielding heads on any given toss of 1/2 and when (n) such coins are tossed, the probability that all the coins land on

mental to science is the idea that regularities may be discovered even in processes with random or chance elements.⁷⁹ Any observation of regularity in complex systems—any model or “law” of nature—depends on statistical theorems that describe the behavior of large samples of objects.⁸⁰ Statistical concepts such as normal distribution,⁸¹ significance levels,⁸² and power⁸³ are all basic to science. Finally, false positives are disfavored in all scientific fields; that is, scientists must

heads is $(1/2)$ to the (n) th power. See *id.* at 114-15. Even after 10 heads have appeared in a row, however, the probability of a tail grows no larger. See *id.* at 31. The coin cannot change its configuration probability since it has no memory.

Equally important is the concept of a random variable. A random variable is a variable whose value is “determined by the outcome of an experiment,” *id.* at 172, such as the number of heads when three coins are tossed, see *id.* at 171-72. For more detailed definitions and explanations of these fundamental statistical concepts, see generally *id.*

⁷⁹ See Mosteller et al., *supra* note 78, at 7-8 (discussing empirical study of variability and demonstrating by experiment with deck of cards that one can measure expected frequency of outcome); Suppes, *supra* note 76, at 100 (discussing probability theory).

⁸⁰ See Finkelstein & Levin, *supra* note 78, at 107 (“[A]ny regularity or ‘law’ of nature observed at ordinary scales is itself a consequence of statistical theorems that describe the aggregate behavior of large samples of objects.”).

⁸¹ Normal distribution is the theoretical probability distribution sometimes referred to as Gaussian distribution or, more commonly, the bell curve. See Colton, *supra* note 14, at 81-82. As one scientist explained, “[t]his distribution is the most common in nature and is such that two-thirds of all values are within one standard deviation . . . of the mean (or average value for the entire population) and 95% are within 1.96 standard deviations of the mean.” Shayne C. Gad, *Statistics and Experimental Design for Toxicologists* 6 (3d ed. 1999). Generally, results of an experiment will not be identical, but will cluster around a particular mean value. Scientists then make inferences based upon that mean value and the amounts that the results deviate from that value. See Colton, *supra* note 14, at 81-84.

⁸² Statistical significance is set by convention at a level of significance, or p-value, of .05 (which corresponds to a confidence level of 95%). See Shayne C. Gad & Carrol S. Weil, *Statistics for Toxicologists, in Principles and Methods of Toxicology* 221, 223 (A. Wallace Hayes ed., 3d ed. 1994) [hereinafter *Principles and Methods*]. Statistical significance tests keep the scientist from asserting a positive effect when the effect may actually be due to chance. See David Ozonoff & Leslie I. Boden, *Truth & Consequences: Health Agency Responses to Environmental Health Problems*, *Sci. Tech. & Hum. Values*, Summer/Fall 1987, at 70, 73-74. Statistical significance describes the probability that the procedure produced the observed effect by chance. Cf. Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 *Jurimetrics* 1, 25 (1995) (discussing confidence intervals). If a test is not statistically significant, it may either be because the results were due to chance or because the test lacked the power to discern a difference between no effect (the “null hypothesis”) and the hypothesized effect. See *id.* at 22.

⁸³ The power of a test refers to the “chance that a . . . test will declare an effect when there is an effect to declare,” or “[m]ore precisely, . . . the probability of [the test] rejecting the null hypothesis when the alternative hypothesis is right.” David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in *Federal Judicial Ctr. Manual*, *supra* note 9, at 331, 381 & n.152. Power increases with the size of the study and with the degree of difference between the alternative hypothesis and the null hypothesis; the more extreme the alternatives, the better the power. See Fienberg et al., *supra* note 82, at 22. Power, therefore, will be an issue for small studies of low effects—precisely those most likely to be proffered in toxic tort cases. Thus, separate studies of small numbers of subjects may not yield statistically significant results simply because each individual test may lack the power to distinguish the null hypothesis of no effect from results that are not extreme. See *id.*

attempt to minimize the probability of claiming an effect when there is none.⁸⁴

Probabilistic assessments also employ assumptions, taking it as given that all knowledge is incomplete.⁸⁵ Background assumptions are pervasive and inevitable;⁸⁶ such assumptions are not unique to statistics, as every branch of science employs them.⁸⁷ No experiment can be performed without using assumptions—but which assumptions are justified? Within a particular theory, or model, scientists agree on a set of basic premises.⁸⁸ These premises are assumptions that have an empirical and consensual basis, and they operate just as rebuttable presumptions work in the law: Under most circumstances, scientists will rely on these background assumptions, but in some circumstances the assumptions will be demonstrably false. In these cases the scientist needs to explain the discrepancy.⁸⁹

⁸⁴ See Colton, *supra* note 14, at 116 (explaining implications of choosing significance level of 5% as chance that, with repeated tests at 5% significance level, on average 95% “will correctly not reject true null hypotheses”); Troyen A. Brennan & Robert F. Carter, *Legal and Scientific Probability of Causation of Cancer and Other Environmental Disease in Individuals*, 10 *J. Health Pol. Pol’y & L.* 33, 72 n.158 (1985) (noting that scientists attempt, whenever possible, to control and eliminate inaccuracies).

⁸⁵ Cf. Stephan F. Lanes, *The Logic of Causal Inference*, in *Causal Inference*, *supra* note 75, at 59, 64-65 (stating that “knowledge is always theoretical” since one cannot “calculate the probability that a theory is true”); see also Frederick Suppe, *The Search for Philosophic Understanding of Scientific Theories*, in *The Structure of Scientific Theories* 3, 107-108 (Frederick Suppe ed., 2d ed. 1977) (noting pervasiveness of assumptions in experimental design and execution). This is the idea behind Lakatos’s revised theory of falsification, in which a series of theories, rather than a single theory in isolation, is subjected to critique and a new series is preferable if it increases the empirical content of the explanation. See Lakatos, *supra* note 58, at 32-35.

⁸⁶ See Naomi Oreskes et al., *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences*, 263 *Science* 641, 641 (1994) (using example of numerical simulation models in earth sciences, and noting that “the observation and measurement of both independent and dependent variables are laden with inferences and assumptions”).

⁸⁷ For a smattering of articles employing assumptions in widely varying fields, see, for example, S. Baumgartner et al., *Geomagnetic Modulation of the ³⁶Cl Flux in the GRIP Ice Core, Greenland*, 279 *Science* 1330 (1998) (geomagnetic force shields); Brush, *supra* note 46 (Newtonian physics); Eric W. G. Diau et al., *Femtosecond Activation of Reactions and the Concept of Nonergodic Molecules*, 279 *Science* 847 (1998) (molecular chemistry); Thomas R. Knutson et al., *Simulated Increase of Hurricane Intensities in a CO₂-Warmed Climate*, 279 *Science* 1018 (1998) (hurricane prediction); Lynn Arthur Steen, *The Science of Patterns*, 240 *Science* 611 (1988) (statistics); see also Alan E. Shapiro, *New Turnings*, 250 *Science* 1600, 1600 (1990) (book review) (stating that scientific revolution was based on conceptual transformation that involved new set of assumptions).

⁸⁸ Collective judgment about which background assumptions are warranted and under what circumstances is not static; this consensus is subject to revision, and this collective judgment forms what Lakatos calls a “research programme.” See Lakatos, *supra* note 58, at 47-52.

⁸⁹ Kuhn points out that sometimes the anomalies in which background assumptions fail are just ignored, but when too many accumulate, the theoretical framework (or research programme, as Lakatos called it) collapses and is replaced by a new framework (or programme). See Kuhn, *supra* note 49, at 81-85.

True verification and validation of theories is impossible, and therefore theories must be evaluated in relative terms. Data (from experiment or from observation), theory, contingencies, and assumptions are juxtaposed to provide an explanation for the way things work.⁹⁰ The operative question is which theories explain the data, not which data establish the theory.⁹¹ Causal inference—like other scientific hypotheses—is a matter of explanation.⁹² This is the foundation of probabilistic reasoning, and judges must understand this to make scientific admissibility determinations properly.

2. *Falsifiability, Criticism, and Rationality*

Karl Popper's philosophy of science has had a marked impact on the way scientists think about what they do.⁹³ Although Popper certainly is not the most recent philosopher of science, nor is he uncontroversial, subsequent philosophers of science have had to respond to his ideas.⁹⁴ Popper's thesis argues that, among other things, a valid theory must be falsifiable.⁹⁵ Popper defined falsifiability as comprising three concepts: testability, openness to critique, and rationality.⁹⁶ These ideas remain central to scientific argument.⁹⁷ Therefore, they

⁹⁰ See Knorr-Cetina, *supra* note 47, at 21 (stating that observation and interpretation, and fit between them, are concerns basic to scientific reasoning); Oreskes et al., *supra* note 86, at 642 (explaining that most that can be said of scientific model is that it "does not contain known or detectable flaws and is internally consistent" and is therefore "valid," in sense of being "legitimate," rather than "true"). See generally Suppes, *supra* note 76.

⁹¹ See Norwood Russell Hanson, *Patterns of Discovery* 64 (1958) (discussing connection of theories of cause and effect).

⁹² See Lanes, *supra* note 85, at 66 (advancing Popperian view of causal inference).

⁹³ For an historical account of Popper's influence, see generally Ian Hacking, *Representing and Intervening* (1983); Suppe, *supra* note 85.

⁹⁴ For a number of prominent scientists who have responded to Popper's ideas, see, e.g., Paul Feyerabend, *Against Method* 93-98 (1978) (disagreeing with Popper by advocating positive role of ad hoc hypotheses in development of scientific theories); Howson & Urbach, *supra* note 72, at 224 (acknowledging pre-eminence of Popper's philosophy while attempting to refute it); Lakatos, *supra* note 58, at 8-9 (discussing Popper's theory of scientific revolution and Kuhn's criticism of it).

⁹⁵ Many scientists ignore Popper's principle in practice. See Beecher-Monas & Garcia-Rill, *supra* note 5, at 256 (noting that "[i]n practice, however, many scientists strive to prove their theories correct (not incorrect)" and thus "[w]hile non-scientists believe that scientists follow the principle of falsifiability, in reality many do not"). However, the fact that scientists think they are performing experiments to confirm a prediction, rather than to falsify it, does not make their belief correct. See, e.g., O'Connor, *supra* note 42, at 274.

⁹⁶ See Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* 37 (5th ed. 1989) ("[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability."). For further explanation of these concepts, see generally Popper, *supra* note 36. For a discussion of Popperian notions of falsifiability, diversity, and rationalism, see Beecher-Monas & Garcia-Rill, *supra* note 5, at 256.

⁹⁷ See Oreskes et al., *supra* note 86, at 645 n.36 (demonstrating continued centrality of Popper's philosophy by challenging use of geological modeling in Popperian terms). Even Popper's critics, who challenge his ideas as insufficient to explain what really goes on in

are useful ideas for judges faced with the task of assessing scientific validity.⁹⁸

As the Supreme Court acknowledged in *Daubert*, testing hypotheses is a key factor in determining “good” science.⁹⁹ Affirming the importance of empirical content to scientific theory, testability suggests that what makes a hypothesis scientific is that it can be corroborated (provisionally) or falsified by observation and experiment.¹⁰⁰ The greater the empirical content of a theory, the stronger the theory, because it is more open to testing and more easily falsifiable.¹⁰¹

As a corollary to testability or falsification, the hypothesis being tested must have precise logical consequences that are incompatible with alternative hypotheses. The articulation of what the testing means is also critically important.¹⁰² Popper explained that the scientist’s task is continually to criticize and reassess scientific theories in light of new theories and data.¹⁰³ The importance of a control group (for experimental studies) or a null hypothesis (for observational studies) is that by such means a researcher exposes the chosen hypothesis to the possibility of falsification. If the data, gathered with proper

science, are forced to acknowledge the force of his philosophy. See, e.g., Alfred Jules Ayer, *Language, Truth and Logic* 38 (2d ed. 1946) (critiquing falsifiability on basis that falsifiability, like confirmation, presupposes certain conditions); see also *supra* note 94.

⁹⁸ As Lakatos explained, theories cannot be appraised without some presuppositions regarding the growth of knowledge. See Lakatos, *supra* note 75, at 159.

⁹⁹ See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 593 (1993) (citing Popper for importance of falsifiability in determining whether “a theory of technique is scientific knowledge”); cf. Gross, *supra* note 59, at 32 (explaining that “commitment in science. . . rests ultimately on agreed-upon procedures”).

¹⁰⁰ See Francisco J. Ayala, *Biology as an Autonomous Science*, 56 *Am. Sci.* 207, 207 (1968) (stating that testability requires that “scientific explanations have precise logical consequences which can be verified or falsified by observation and experiment”). A theory can be corroborated at any given point in time only provisionally; further testing may reveal hitherto undiscovered flaws that refute it. See Popper, *supra* note 36, at 275.

¹⁰¹ See Popper, *supra* note 36, at 121 (observing that rule that prefers theories “which can be most severely tested . . . [is] equivalent to a rule favouring theories with the highest possible empirical content”). Thus, “[s]cience advances toward truth (though never arriving at certainty) by a combination of bold conjecture and severe criticism.” Gieryn, *supra* note 51, at 395; see also Faigman, *supra* note 49, at 1018 (“[T]he fact that scientific theories are vulnerable to falsification imparts a strength stemming from having taken the risk of refutation.”).

¹⁰² See Lakatos, *supra* note 75, at 109-10 (explaining that articulating meaning is fundamental precondition for appraisal).

¹⁰³ See Popper, *supra* note 36, at 275 (explaining that appraisal of statements is relative to information available at particular point in time). In addition, the effects of secondary variables—variables other than the ones examined by the test—either must be controlled, known (as in an experiment), or explicitly assumed (as in observational studies of human populations). For a descriptive distinction between experiment, quasi-experiment, and observational study, see Fienberg et al., *supra* note 82, at 15-17.

controls, do not refute the proposed hypothesis, it may be (provisionally) valid.¹⁰⁴

The concept of testability has been challenged on several fronts.¹⁰⁵ Testability alone ignores the problem of identifying underlying assumptions.¹⁰⁶ Testability assumes a laboratory science paradigm of methodology; the idea must be stretched to encompass epidemiology, for example, with its reliance on statistical inference rather than controlled experiments.¹⁰⁷ Even fields outside the laboratory paradigm, however, have criteria for validity that need to be met, and thus their own form of testability. But testability does not stand alone as the sole criterion for distinguishing science from pseudoscience.¹⁰⁸

In addition to testability, Popper believed that valid science requires theories to be open to information from and critique by the entire scientific community.¹⁰⁹ Instead of relying on a narrow focus, Popper believed that the interaction of various scientific disciplines was important to the growth of knowledge.¹¹⁰ He thought that scientific inquiry could only benefit by listening to interdisciplinary

¹⁰⁴ As one scientist explains, “[t]he practice of scientific method is the persistent critique of arguments in the light of tried canons for judging the reliability of the procedures by which evidential data are obtained, and for assessing the probative force of the evidence on which conclusions are based.” Ernest Nagel, *The Structure of Science* 13 (1961).

¹⁰⁵ For a number of such challenges, see, e.g., Lakatos, *supra* note 58, at 32-33 (arguing that Popper’s proposed asymmetry between falsification and corroboration is much less significant than he thought); Conley & Peterson, *supra* note 9, at 1201-02 (presenting criticisms of Supreme Court’s reliance on testability as ultimate criterion for what makes something scientific knowledge). In fact, some commentators have argued that using falsifiability as a criterion for good science would render much of modern science unscientific. See, e.g., Alan Chalmers, *Science and Its Fabrication* 18 (1990) (contending that if criterion of falsifiability is “formulated sufficiently strongly to have some force, then physics would fail to qualify as a science”). What is important, the critique suggests, is that any new theory must explain the partial success of its predecessor and also something more. See Lakatos, *supra* note 58, at 142.

¹⁰⁶ See Lakatos, *supra* note 58, at 1-7.

¹⁰⁷ Testability also may not be an adequate notion for psychology and other social sciences where retrospective observational studies, rather than controlled experimentation, are the norm. Empirical data, statistical description, openness to critique, and coherence with the underlying theory are also indicia of validity in these fields. The Supreme Court made this point in *Kumho Tire Co. v. Carmichael*, 526 U.S. 136 (1999), where it emphasized that *Daubert* analysis must be applied even to “soft science” fields such as engineering and psychology. See *id.* at 147.

¹⁰⁸ For a discussion of the critique that, although testability is necessary, it is not sufficient to distinguish good science from bad science (pseudoscience), see Lakatos, *supra* note 58, at 1-7.

¹⁰⁹ See Popper, *supra* note 36, at 44 & n.1; Karl Popper, *Normal Science and Its Dangers*, in *Criticism and the Growth of Knowledge* 51, 57 (Imre Lakatos & Alan Musgrave eds., 1970) (arguing that science progresses through “critical comparison of the competing frameworks”).

¹¹⁰ See Popper, *supra* note 36, at 44 & n.1.

voices.¹¹¹ The more that different fields of science interact to learn from and critique each other's ideas, the more we will know about the world.¹¹² Using studies from many disciplines thus forms an important facet of good scientific argument.¹¹³

The force of rationality is a third underlying tenet of Popper's philosophy of science.¹¹⁴ Popper believed that rationality consists of "stating one's problem clearly and . . . examining its various proposed solutions *critically*."¹¹⁵ Popper thus equated rationality with a critical attitude.¹¹⁶ Rationality consists of comparing ideas and has little to do with belief.¹¹⁷ A background in science is not required to understand scientific ideas.¹¹⁸ Rather, scientific ideas ought to be tested in a public forum by experts from many different fields.

For Popper, rationality encompasses testing the internal consistency of the theory or hypothesis, testing the consistency with other theories, and testing the consistency with the experimental data.¹¹⁹ Indeed, Popper distinguished science from myth by science's falsifiability in which Popper included its testability, openness to interdisciplinary critique, and rationality.¹²⁰ However, it is the explanatory power of increased empirical content that makes one falsifiable theory superior to another.¹²¹

¹¹¹ See *id.* at 399.

¹¹² An important example of this interdisciplinarity is the rapid spread and application of chaos theory and principles of nonlinearity to a broad spectrum of scientific disciplines, including chemistry, physics, and ecology, among others. See Celso Grebogi et al., *Chaos, Strange Attractors, and Fractal Basin Boundaries in Nonlinear Dynamics*, 238 *Science* 632, 632 (1987) (noting "explosive growth" in field of chaotic dynamics of dissipative systems, and its many applications "across a broad spectrum of scientific disciplines").

¹¹³ See Oreskes et al., *supra* note 86, at 645 n.38 (noting that "diversity in confirmation helps to explain why it is important to test a model in a wide variety of circumstances—including those that may appear quite different from the expected circumstances").

¹¹⁴ See Popper, *supra* note 36, at 16-17 (describing "rational discussion" as basic method of natural sciences and philosophy).

¹¹⁵ *Id.* at 16.

¹¹⁶ See *id.*

¹¹⁷ See, e.g., Lakatos, *supra* note 75, at 196 n.10.

¹¹⁸ Popper differed radically from elitists such as Polanyi and Feyerabend, who stressed the notion that only a privileged elite can understand science. See Lakatos, *supra* note 58, at 91 (critiquing Polanyi who regarded elaboration of rational standards as hopeless enterprise, in contrast to Popper and Lakatos who developed standards for rationality).

¹¹⁹ See Popper, *supra* note 36, at 40.

¹²⁰ See *id.* at 16 (noting force of rationality); *id.* at 40 (discussing falsifiability as criterion for demarcation); *id.* at 44 n.*1 (noting importance of interdisciplinary critique). Thus, a theory is not better for Popper just because a scientific elite prefers it, but because the ideas have withstood severe criticism and offer a better, more persuasive, rational explanation that is higher in empirical content than rival theories. Cf. Lakatos, *supra* note 75, at 228-42 (arguing for clear theses in science "where logic can assist criticism and help to appraise the growth of knowledge").

¹²¹ See Popper, *supra* note 36, at 277 & n.*1 (discussing explanatory power of scientific theories).

3. *Explanatory Power*

The explanatory power of a scientific theory is its ability to explain what happened. A testable, critiquable, and rational scientific theory is still useless, unless it can tell us why the results turned out the way they did, and what the new theory explains that the previous one did not.¹²² Empirical support for this explanation may take many forms.¹²³ Thus, scientists gather all the information available in support of a plausible theory. Scientific explanation consists of juxtaposition of theory and data, that is, the ability to take data and turn them into a valid scientific argument.¹²⁴ What makes one theory stronger than another, or why one theory should supplant a prior theory, is a function of its explanatory power.

The explanatory power of science builds on metaphor.¹²⁵ A scientific model is a metaphor, an analogy to something more familiar.¹²⁶ The better the model, the more it will take into account all of the data tested or observed.¹²⁷ The most common scientific metaphors (or models) in current use are the pump and the computer, and many disparate systems are described in terms of these metaphors.¹²⁸ The heart, for example, is commonly described as a pump,¹²⁹ and the brain

¹²² See Lakatos, *supra* note 58, at 32 (“[A] scientific theory T is *falsified* if and only if another theory T’ has been proposed with the following characteristics: (1) T’ has excess empirical content over T . . . (2) T’ explains the previous success of T . . . and (3) some of the excess content of T’ is corroborated.”).

¹²³ As noted in the Federal Judicial Center Manual, “[m]ultiple avenues of deductive reasoning based on research data lead to scientific acceptance of causation in any field” Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, in Federal Judicial Ctr. Manual, *supra* note 9, at 181, 212 (acknowledging difficulty of quantitatively describing process of inferring causation from multiple studies).

¹²⁴ See generally Hacking, *supra* note 93 (examining relationship between data, observation, and theory).

¹²⁵ See *supra* note 65 and accompanying text (explaining use of models and metaphors in science).

¹²⁶ See Gross, *supra* note 59, at 80-81 (arguing that metaphor seems to have indispensable role in science, even though use of metaphor seems inconsistent with mature scientific knowledge).

¹²⁷ See Ayala, *supra* note 100, at 207 (stating that “distinctive characteristic of science” is “that it strives to provide explanations of why the observed events do in fact occur”); Oreskes et al., *supra* note 86, at 644 n.19 (observing that “scientists routinely modify their models to fit recalcitrant data”).

¹²⁸ See, e.g., *id.* at 81 (“[A] unit of a living cell is a pump, and a biochemical process is a cascade that embodies a positive feedback mechanism.” (internal quotation marks omitted)).

¹²⁹ See, e.g., Marcia Barinaga, How Much Pain for Cardiac Gain?, 276 Science 1324, 1325 (1997) (referring to heart as pump).

is described as a computer.¹³⁰ These metaphors are useful but, to some degree, inaccurate.¹³¹

Explanatory power encompasses the idea of reliability, which in science refers to the reproducibility of results.¹³² The explanatory power of a theory, however, is more than reliability. Explanatory power is what distinguishes a good experiment from one that merely can be reproduced. Just because results are reproducible, however, does not mean that they are correct. On the other hand, the irreproducibility of results suggests that something is wrong with the methodology.¹³³

It is important to remember, however, that although the explanatory power, or persuasiveness, of a theory can make it superior to competing theories, it is impossible to say whether a given theory is "true."¹³⁴ It is superior if it offers a better metaphor, and if it accounts for most, if not all, of the known data in a more persuasive way than other theories.¹³⁵ In addition, the theory must be capable of accommodating newly discovered information.¹³⁶

It is around these basic principles of testability, criticism, rationality, and explanatory power—principles inherent in the scientific method—that any heuristic must be framed if it is to be useful to judges in assessing the admissibility of scientific arguments.

II

THE HEURISTIC: A GUIDE FOR JUDGES IN DECIDING THE ADMISSIBILITY OF SCIENTIFIC EVIDENCE

A judge confronted with a scientific admissibility determination has to decide whether the scientific expert is testifying about valid scientific knowledge, which *Daubert* explained as knowledge grounded

¹³⁰ See, e.g., Harry Levin, *Successions in Psychology*, 236 *Science* 1683, 1683 (1987) (book review) (describing computer as metaphor for human information processing).

¹³¹ See Beecher-Monas & Garcia-Rill, *supra* note 5, at 256 (noting that inescapable differences between model and actual process, along with "intrinsic problems and hidden assumptions," can render metaphors slightly inaccurate).

¹³² See Kaye & Freedman, *supra* note 83, at 341.

¹³³ For example, the irreproducibility of the cold nuclear fusion experiments tipped off the fraudulent nature of the experiments. For a discussion of the "discovery" of cold nuclear fusion and the repeated attempts to prove its existence, see generally Gary Taubes, *Bad Science: The Short Life and Weird Times of Cold Fusion* (1993).

¹³⁴ See Beecher-Monas & Garcia-Rill, *supra* note 5, at 254-55 ("[W]e only can be definite about the superiority of one theory over another, not about its absolute truth.").

¹³⁵ Lakatos explained that "[a] hypothesis, however novel in its intuitive aspects, will not be allowed to be proposed, unless it has novel empirical content in excess of its predecessor." Lakatos, *supra* note 58, at 142.

¹³⁶ See Popper, *supra* note 36, at 279 (stating that accommodating newly discovered evidence is hallmark of valid theorizing).

in and derived by the scientific method.¹³⁷ Scientific method, insofar as it exists, consists of a common discourse about probabilistic assessment, testability, criticism, rationality, and explanation. How can these themes be teased into a useful heuristic to solve the validity puzzle?

There are five basic factors a judge should consider in making the decision to admit expert testimony into evidence. First, the judge must identify and understand the underlying theory and hypothesis. Second, in order to determine whether the theory is supportable, the judge must examine all of the available information—human studies, animal studies, cellular studies, and chemical structure—in concert.¹³⁸ Third, where there are information gaps (which are inevitable), judges should fill them with scientifically justifiable default assumptions. Fourth, the judge should conduct an inquiry into the methodology (including the laboratory or observational methods as well as statistical methodology) and assess whether it conforms to generally acceptable practices in the field. Finally, the judge must put all this information together in such a way as to make a probabilistic assessment of the strength of the links between theory, assumptions, methodology, and the conclusion the expert espouses. This Part explains each of these five steps in detail.

A. What's the Theory? Examine the Explanatory Power of Theory and Hypothesis

Expert testimony is proffered to support a hypothesis. Underlying the hypothesis is a more general theory. Theories are the starting point for scientific analysis.¹³⁹ The difference between theory and hypothesis is a matter of degree: A theory has undergone more testing and refinement (and therefore gained more acceptability) than has a hypothesis.¹⁴⁰ Judges cannot decide whether a scientific theory or hypothesis is correct.¹⁴¹ Neither can scientists. Instead, the most anyone can decide in making a decision about whether a given theory or hypothesis is “valid” is whether the theory is supported adequately

¹³⁷ See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590 (1993).

¹³⁸ In this vein, it is unscientific to expect a single study to uphold an entire explanation. See *infra* notes 143-48 and accompanying text (arguing that courts should consider studies in combination with one another to determine whether all proffered testimony and evidence support expert opinions).

¹³⁹ See Brennan, *supra* note 68, at 483 (“[S]cientists understand that theories define uncertainty and provide the basis for hypothesis building.”).

¹⁴⁰ See Popper, *supra* note 36, at 121 (stating that if hypothesis repeatedly withstands testing it may become accepted as theory).

¹⁴¹ See *supra* notes 74-76 and accompanying text (explaining that theories cannot be proven true, or even probable).

enough by facts and logic to be reliable, either as evidence at trial or—in the case of a scientist—as a basis for further experimentation. Understanding science as a process of idea construction rather than mere description makes it possible for a judge to examine the logic of the ideas about which the expert proposes to testify and how those ideas are rationally related to what they are intended to show.

Knowing what the underlying theory is and whether it makes sense—both on a general level and as it applies to the specific case—is the first step in this process. In a toxic tort case, for example, the underlying theory on the most general level is that exposure to chemicals can cause illness or birth defects; less broadly, that a particular chemical can cause specific effects; and at the level of specific causation, that certain chemicals to which plaintiffs were exposed caused plaintiffs' particular disease or birth defects.¹⁴² In a criminal case where identification is the issue, the theory is that each individual has unique identifying characteristics (DNA, fingerprints, handwriting, etc.) shared by relatively few other people, that these characteristics can be measured, and that the frequency with which these characteristics appear in the general population is quantifiable and therefore helps identify the defendant.

Identifying the theory is deceptively simple, although courts rarely stop to explain the theory that underlies the expert's testimony. Assessing the theory's validity—that is, determining whether each of the theory's parts fit together to make a coherent whole—is more difficult, and courts routinely get it wrong. Yet the court must identify and understand the underlying theory before moving on to evaluate the data, assumptions, methodology, and the strength of the links between them.

A good theory accommodates most or all of the available evidence. It explains data that appear to refute it. Examining all the available data is therefore key to deciding whether the theory explains the phenomenon well enough to be admissible.

B. Examine All of the Available Data

Many courts—including the Supreme Court in *Joiner*—mistakenly require that each study proffered by an expert support the entirety of the expert's hypothesis in order for that expert's testimony to

¹⁴² Civil courts refer to the broad theory level as general causation; the more specific theory level is referred to as specific causation. See, e.g., *Cavallo v. Star Enter.*, 892 F. Supp. 756, 771 n.34 (E.D. Va. 1995) (explaining difference between general and specific causation).

be admissible.¹⁴³ This notion results from a fundamental misconception about the scientific process. No single study should be expected to establish the validity of a theory on its own, nor should the supporting data be limited to one particular field of inquiry.¹⁴⁴ Not only are many steps and sources of data necessary to build a theory, but each step itself builds on a theory comprising many steps, and so on. Moreover, experimental studies using different methodologies (such as in vivo animal studies, in vitro tissue culture studies, and structure-activity relationship studies)¹⁴⁵ may produce inferential synergy.¹⁴⁶ That is, taken together such studies may have far greater inferential force than if considered separately.¹⁴⁷ Of course, not all evidence will corroborate the expert's hypothesis, and the expert should be able to explain contradictory or conflicting data.

In *Joiner*, Justice Stevens properly expressed annoyance that the parties had presented a record devoid of the studies and data on which the experts relied.¹⁴⁸ An assessment of the studies' scientific validity could not be undertaken without access to the underlying

¹⁴³ The Supreme Court in *Joiner* made this mistake when, rather than assessing how the various studies in conjunction supported the experts' hypothesis, the majority seemingly approved the district court's method of examining the studies seriatim as well as the court's conclusion that none was individually sufficient to justify the experts' causation conclusion. See *General Elec. Co. v. Joiner*, 522 U.S. 136, 146-47 (1997). The Court found that it is within the trial judge's discretion to determine whether proffered studies, "whether individually or in combination," are sufficient. See *id.*

¹⁴⁴ Stephen E. Fienberg and coauthors emphasize the scientific merit of combining different studies. They comment:

Complex scientific inferences are rarely the result of a single statistical inference, but are based on disparate pieces of scientific evidence of different perspectives in different contexts. For example, a cause-and-effect relationship may be inferred from a consistent association in several epidemiological studies, a relationship between exposure or dose and response in toxicological experiments with animals, and a causal mechanism consistent with some biological theory.

Fienberg et al., *supra* note 82, at 13; see also *supra* notes 109-13 and accompanying text.

¹⁴⁵ See *infra* note 149 and accompanying text (discussing value of drawing biological explanations from studies in various fields of science).

¹⁴⁶ See David A. Schum, *The Evidential Foundations of Probabilistic Reasoning* 282-83 (1994) (discussing ways in which different items of evidence can be "synergistic" such that they may enhance or reduce force of other evidence).

¹⁴⁷ See *id.* at 283 (emphasizing that "one item of evidence can act to *enhance* the inferential force of another" while noting that "there are other situations in which one item of evidence can act to reduce the inferential force of another item" (emphasis added)).

¹⁴⁸ See *Joiner*, 522 U.S. at 151 (Stevens, J., concurring in part and dissenting in part). Justice Stevens argued that the district court erroneously excluded studies piecemeal rather than weighing all the available scientific evidence. See *id.* at 152-53. Assessing the facts of *Joiner*, Justice Stevens concluded that had the district court permitted the proffered studies to be used in conjunction with each other, "an expert could reasonably have concluded that the study of workers at an Italian capacitor plant, coupled with data from Monsanto's study and other studies, raises an inference that PCB's promote lung cancer." *Id.* at 154.

data. What kind of data should judges be looking for? In the best of all possible worlds, a sound biological explanation would be drawn from many well-performed studies of humans, animals, cells, general metabolic processes, and chemical structures.¹⁴⁹ Each of these kinds of studies provides different information contributing to an understanding and recognition of the importance of the other studies.

Many courts refuse to apply this kind of wide-ranging, interdisciplinary analysis, mistakenly believing that it is only a “methodology result[ing] from the preventive perspective that the agencies adopt in order to reduce public exposure to harmful substances,” and not a useful means of establishing the reliability necessary for judicial admissibility.¹⁵⁰ However, this multidisciplinary approach is used not

¹⁴⁹ See, e.g., Ronan O’Rahilly & Fabiola Müller, *Human Embryology & Teratology* 8-9 (2d ed. 1996) (explaining importance of animal studies to understanding human development); James L. Schardein, *Chemically Induced Birth Defects* 25-37 (3d ed. 2000) (discussing use of structure-activity studies, pharmacokinetics, animal models, and in vitro methods for studying birth defects); Barbara D. Beck et al., *The Use of Toxicology in the Regulatory Process*, in *Principles and Methods*, supra note 82, at 19, 25 (observing in context of regulatory decisionmaking that “[t]he rational approach is to examine all sources of information in the evaluation of toxic chemicals”); Christopher P. Chengelis & Shayne Cox Gad, *Introduction to Animal Models in Toxicology* 1, 1 (Shayne Cox Gad & Christopher P. Chengelis eds., 1992) (“[E]xperiments in animals have provided the necessary building blocks that permitted the explosive growth of medical and biological knowledge in the later half of the twentieth century.”); Ernest Hodgson, *Introduction to A Textbook of Modern Toxicology* 1, 6-7 (Ernest Hodgson & Patricia E. Levi eds., 2d ed. 1997) (explaining interrelatedness of toxicology with other sciences).

The EPA uses this broad-based or “weight of the evidence” approach to assess the risk of cancer or birth defects presented by various chemicals. See EPA Proposed Guidelines, supra note 10, at 17,961 (explaining that “decisions come from considering all of the evidence”). This same approach is used by the World Health Organization’s International Agency for Research on Cancer (IARC) and by the National Toxicology Program (NTP) of the National Institute of Environmental Health Science (NIEHS). See Bert Black, *Post-Daubert and Joiner Caselaw: The Good, The Bad, and the Ugly*, 1998 A.L.I.-A.B.A. Course of Study: Products Liability 147, 154 (discussing *Allen v. Pennsylvania Eng’g Corp.*, 102 F.3d 194, 196 (5th Cir. 1996), and noting that IARC, OSHA, and EPA all follow “weight of the evidence” approach); Cranor et al., supra note 39, at 69-70 (stating that NTP and IARC use same criteria for evaluating carcinogens). In this respect, the agencies have it right.

¹⁵⁰ *Allen*, 102 F.3d at 198. In *Allen*, the plaintiff’s experts sought to testify that the plaintiff’s exposure to ethylene oxide had caused his brain cancer. The experts relied on imperfect epidemiological studies, animal studies, and cell biology data, none of which standing alone was sufficient to explain the plaintiff’s illness. See *id.* The Fifth Circuit took notice of the recognized connection between ethylene oxide exposure and other kinds of cancers, the “suggestive” link between the chemical and brain cancer demonstrated by the proffered epidemiological studies, and considered both proffered animal studies showing brain tumors developed in rats after exposure to the chemical and cell biology studies showing the chemical at issue to have “mutagenic and genotoxic capabilities.” See *id.* at 197-98. Despite all of this, the court found that the “paucity of epidemiological evidence, the unreliability of animal studies, and the inconclusiveness of cell biology combine to undercut the expert testimony.” *Id.* at 198. The Fifth Circuit therefore held that the district court properly excluded the experts’ testimony. See *id.*

just in preventive agency determinations, but is applied widely.¹⁵¹ The broad-based, multistudy approach is the most widely used methodology by the scientific community and has firm scientific support.¹⁵²

Although an expert's unexplained adoption of agency risk assessments should be viewed critically by trial judges because they may be based on underlying default assumptions meant to protect the public, that does not mean the entire methodology should be categorically dismissed. Such a rejection fails to recognize that most agency determinations are based on scientific probability assessments rather than on worst-case scenarios, and therefore may assess the environmental risk realistically.¹⁵³

Most importantly, assessing the cumulative force of all the available information is consistent with modern research methods.¹⁵⁴ Science accommodates change. Judges too should accommodate changes in research. Although human studies may be the most relevant in assessing the effect of a given chemical on human beings, animal studies and information about the physical, structural, and chemical properties of a suspected agent also can yield important information about its risk to humans.¹⁵⁵

¹⁵¹ In focusing on the regulatory aspect of the "weight of the evidence" approach, courts are confusing risk *assessment* with risk *management*. That is, the EPA may make a risk assessment that a chemical has only a slight possibility of being carcinogenic, for example, and nevertheless decide as a matter of risk management to regulate it to protect public health so that people are exposed at only very low levels. See EPA Proposed Guidelines, *supra* note 10, at 17,965.

¹⁵² See, e.g., Goldstein & Henifin, *supra* note 123, at 186; Interdisciplinary Panel on Carcinogenicity, *Criteria for Evidence of Chemical Carcinogenicity*, 225 *Science* 682, 686 (1984) ("[E]ven in the instances where data are available from humans, the data must be supplemented with information from other sources before a conclusion can be reached."). Although EPA risk assessments should not be adopted necessarily as dispositive in toxic tort litigation, there is nothing inherently biased about the methodology used in reaching the risk characterization. But see Ellen Relkin, *The Sword or the Shield: Use of Governmental Regulations, Exposure Standards and Toxicological Data in Toxic Tort Litigation*, 6 *Dick. J. Env'tl. L. & Pol'y* 1, 7-9 (1997) (cautioning against wholesale adoption of regulators' risk assessment models).

¹⁵³ Indeed, the EPA's Proposed Guidelines explain that while initial assessments may be "worst case" in their default assumptions, subsequent risk assessments based on more complete information will replace the initial assessments as more data become available. See EPA Proposed Guidelines, *supra* note 10, at 17,965.

¹⁵⁴ See *id.* at 17,961 (stating that examination of all available information is consistent with "the growing sophistication of research methods, particularly in their ability to reveal the modes of action of carcinogenic agents at cellular and subcellular levels as well as toxicokinetic and metabolic processes").

¹⁵⁵ See *id.* at 17,977 (approving use of and finding great relevance in animal studies and studies of "physical, chemical, and structural properties of an agent"). Physicochemical properties of suspected toxic chemicals include how a particular chemical is absorbed by the body, how it is distributed and degraded in the body, and how it reacts with body cells. See *id.* In extrapolating the results of animal studies to humans, and to studies of absorption, distribution, biotransformation, and excretion, one can compare the effects of a par-

The more complete the picture presented by these various studies, the better the judge can assess the expert's conclusions.¹⁵⁶ But all pictures are incomplete. Deciding whether conflicting data and gaps in the information provided by the studies weaken the expert's position or prove fatal to the expert's proffered testimony depends upon a careful evaluation of these uncertainties.¹⁵⁷ In litigation, this means that the party offering scientific evidence (usually the plaintiff in toxic tort actions, or the prosecutor in criminal cases) will have to show how the individual studies on which the expert relies link together analytically to support the inferences the expert wishes the court to draw.¹⁵⁸ But it also means that experts must be permitted to proffer studies other than human studies in support of their hypotheses to fill the gaps in scientific knowledge of human reactions.

Facts alone, even scientific facts, are not knowledge. They become scientific knowledge only in conjunction with coherent answers to the following questions: What range of facts deserves investigation? What is the proper way to investigate them? And what do the results of the investigation mean?¹⁵⁹ Linking the data to the hypothesis requires answering all three of these questions.

Even the best testing process (or observational study) depends, for its scientific validity, on how well (i.e., how logically) the data is analyzed.¹⁶⁰ A scientist needs not only to justify the choice of experi-

ticular chemical among species and determine the implications of animal response for human hazard assessment. See, e.g., *id.* at 17,980 (outlining factors to consider in extrapolating results of animal studies to humans).

¹⁵⁶ See *id.* at 17,980 (noting importance of "complete . . . data package" and "data-rich" source of information).

¹⁵⁷ See *id.*; see also *id.* at 17,981 (noting importance of "a collective evaluation of all pertinent information so that the full impact of biological plausibility and coherence are adequately considered").

¹⁵⁸ See, for example, *Rosen v. Ciba-Geigy Corp.*, 78 F.3d 316, 319 (7th Cir. 1996), where Judge Posner castigated the expert for attempting to infer precipitation of the plaintiff's heart attack from a nicotine patch manufactured by the defendant. The expert based his conclusion on long-term animal studies showing that nicotine can contribute to the formation of plaque over time. Yet as Judge Posner correctly pointed out, the expert had failed to distinguish between long-term studies and the short-term effects of wearing a nicotine patch for three days. In other words, the expert failed to provide a link between the long-term studies and the short-term effects suffered by the plaintiff. See *id.* Thus, the studies did not combine analytically to support the plaintiff's theory.

¹⁵⁹ See Gross, *supra* note 59, at 4 (stating that "scientific knowledge consists of the current answers" to these three questions, and arguing that scientific facts are meaningless without structure provided by answers to these fundamental questions).

¹⁶⁰ Cf. Gad, *supra* note 81, at 26-29 (explaining principles of experimental design and analogizing experimental design to logic flow analysis); Polanyi, *supra* note 45, at 161 (explaining importance of analysis of facts); Brennan, *supra* note 68, at 483 (discussing importance of theory in empirical studies).

mental design,¹⁶¹ but to provide models of data,¹⁶² and to describe how well the data actually fit the expected theoretical model.¹⁶³ The theory of the experiment permeates all these considerations, and the theory may not be assessed adequately without understanding the assumptions upon which it rests.

C. *Use Assumptions to Fill the Gaps*

The scientific process often leads to the phenomenon of “battling experts” both inside and outside the courtroom. Yet battling experts may both offer valid hypotheses. Different approaches to data analysis may lead to radically different conclusions, depending on the researcher’s underlying assumptions and strategies.¹⁶⁴ Even in a controlled, randomized study, two investigators from the same team, assessing the same data, can reach opposite conclusions.¹⁶⁵ It is “often not that one analyst is right and the other wrong[,] but that different assumptions, implemented through different analytical strategies, can produce conflicting results.”¹⁶⁶ These (often unstated) assumptions

¹⁶¹ Issues with respect to choice of design include the number of trials and choice of experimental parameters. See Gad, *supra* note 81, at 21-31 (explaining principles of experimental design).

¹⁶² A model is the formal description of how observations are produced, incorporating various assumptions implicit in the presentation of the evidence. See Fienberg et al., *supra* note 82, at 2.

¹⁶³ See Suppe, *supra* note 85, at 258-59.

¹⁶⁴ See, e.g., Daryl E. Chubin, *Research Malpractice*, in *Science Off the Pedestal: Social Perspectives on Science and Technology* 95, 99 (Daryl E. Chubin & Ellen W. Chu eds., 1989) (“The history of science is littered with examples of how strong belief—pro or con—can alter perception.”). Data is defined and given significance by the particular scientist as well as by the conceptual structure of the scientific discipline in which the scientist is participating. See Rubin, *supra* note 50, at 1883 (“[S]cientists determine the significance of the data they have . . . generated and defined.”).

¹⁶⁵ This is precisely what happened in the clinical trials of antibiotic treatment for “glue-ear,” a common childhood ailment. See Andrew W. Osterland, *A Child’s Guide to Medical Ethics*, *Fin. World*, Aug. 16, 1994, at 54, 54 (chronicling disagreement between Drs. Erdem Cantekin and Charles Bluestone over whether or not antibiotics are effective in treating “glue-ear”). One of the assumptions at stake in the “glue-ear” controversy between Drs. Cantekin and Bluestone may have resulted (at least according to Dr. Cantekin) from drug company funding of the studies. This funding may have skewed results in favor of the drug. See *id.*

¹⁶⁶ George Davey Smith, *Increasing the Accessibility of Data*, 308 *Brit. Med. J.* 1519, 1519 (int’l ed. 1994) (urging scientists to make data public so others can evaluate assumptions and inferences they made). The court-appointed expert in *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Or. 1996), explained the importance of inference-drawing, observing that two experts validly may draw different inferences from the same data. See Letter from Merwyn R. Greenlick to U.S. District Judge Robert Jones and Oregon Court of Appeals Judge Nely Johnson, reprinted in *Hall*, 947 F. Supp. app. B at 1448 (observing that “both [plaintiffs’ and defendants’ experts’ inferences] are based on scientifically valid data” and that they “have arrived at somewhat different positions as a result of different, but legitimate, interpretations of the results”). The court did not respond to this explanation.

must be analyzed to assess whether the expert's conclusion is a valid interpretation of the studies on which the expert relied.¹⁶⁷

Gaps in scientific knowledge are inevitable; they are not fatal flaws.¹⁶⁸ Making assumptions to fill gaps in scientific knowledge is a normal and necessary part of scientific analysis. As one scientist explained:

All scientific work is incomplete whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.¹⁶⁹

Appropriate inferences in the context of litigation should be based not only on the concerns of scientists, but also on judicial policy concerns. Thus, basic assumptions should "ensur[e] scientific validity, minimiz[e] serious errors in estimating risks, maximiz[e] incentives for [safety] research, creat[e] an orderly and predictable process, and foster[] . . . trustworthiness."¹⁷⁰ Merely rejecting studies as "too speculative," as many courts dismissively deem them, without explaining the basis for rejecting the underlying assumptions, is not enough.¹⁷¹ Intel-

¹⁶⁷ Judges commonly assume in assessing scientific testimony that the expert who relies on a particular study must adopt the study's conclusion. See *infra* notes 385-87 and accompanying text. Such an assumption is mistaken, however, because it is the underlying data from the study—no doubt in conjunction with the underlying data from other studies—which the expert uses to reach a different, but (absent any errors in reasoning) equally valid conclusion. "Science as we know it would not exist if scientists did not defend their theories, by criticizing either the assumptions that go into the design of an experiment that supposedly refutes their theories, or the alternative theories that are proposed as replacements for their theories." Warren Schmaus, *An Analysis of Fraud and Misconduct in Science*, in *American Ass'n for the Advancement of Science*, *supra* note 50, at 87, 89.

¹⁶⁸ Inferences based on general scientific knowledge about phenomena are not only acceptable, but unavoidable. See Committee on the Institutional Means for Assessment of Risks to Pub. Health, National Research Council, *Risk Assessment in the Federal Government: Managing the Process* 11 (1983) [hereinafter *Risk Assessment in the Federal Government*] (discussing "pervasive uncertainty" which necessarily surrounds inferences about risk assessment).

¹⁶⁹ Hill, *supra* note 37, at 300.

¹⁷⁰ Committee on Risk Assessment of Hazardous Air Pollutants, National Research Council, *Science and Judgment in Risk Assessment* 7 (1994) (proposing criteria to consider when establishing guidelines for risk assessment).

¹⁷¹ For cases rejecting studies as too speculative without analysis of assumptions, see, for example, *Moore v. Ashland Chem. Inc.*, 151 F.3d 269, 279 (5th Cir. 1998) (upholding court's rejection of chemical toxicity studies because studies were speculative), cert. denied, 526 U.S. 1064 (1999); *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 197 (5th Cir. 1996) (deeming reliance on animal studies of carcinogenicity of ethylene oxide too speculative); *Frank v. New York*, 972 F. Supp. 130, 136-37 (N.D.N.Y. 1997) (stating that multiple chemical sensitivity studies are too speculative); *Cohen v. Secretary of the Dep't of Health & Human Servs.*, No. 94-0353V, 1998 WL 408784, at *15 (Fed. Cl. July 1, 1998) (excluding plaintiff's expert's testimony interpreting prior studies because speculative); *Gherardi v. Secretary of the Dep't of Health & Human Servs.*, No. 90-1466V, 1997 WL 53449, at *4

lectual due process¹⁷² requires explicit acknowledgment of the assumptions used—or rejected—by the court, and the reasons for the court's determination. The following sections discuss the default assumptions upon which sound judicial validity determinations should be based.

Judges, like scientists, should understand rules of thumb in assessing argument. Categorizing studies according to their persuasiveness (roughly analogous to the ranking of legal authority) simplifies and clarifies decisionmaking. An important caveat, however, is that, as with legal rules of thumb, the scientific hierarchy is crude and is frequently inapplicable and ignored by scientists. Even the "lowest" kinds of studies in the rough hierarchy are still good studies offering valuable information. Thus, although there may be good reasons to disregard the hierarchy, it is important nonetheless to understand the hierarchy and its assumptions.

The following rules of thumb regarding rankings and assumptions, although flawed and not always applicable, are still employed often by scientists and thus should be understood by judges in assessing the validity of scientific studies:

- (1) Statistical methods are critical to making appropriate assumptions, but statistical methods themselves rely on a number of assumptions.
- (2) In studies of human disease, human studies are most valuable, followed by animal studies, in vitro studies, and studies comparing the relationship of an element's chemical structure to the element's effects. However, properly performed and controlled human studies are scarce, and each type of study offers important information.
- (3) The assumption made by identification experts, such as forensic scientists, that all humans have unique characteristics, is problematic in all but a few areas (such as DNA identification techniques).

As previously noted, as with all rules of thumb, there are numerous caveats, exceptions, and countervailing notions to these rankings that may make them inapplicable in particular instances. Thus, although the general hierarchy that follows has some scientific merit and simplifies decisionmaking, it is not necessarily optimal. Things are not really as simple as this hierarchy suggests. Each of the canons has a counterpoint. Each type of methodology has inherent strengths

(Fed. Cl. Jan. 24, 1997) (deeming evidence of tainted vaccine too speculative), *aff'd*, No. 99-5018, 2000 WL 158469 (Fed. Cir. Feb. 10, 2000); *Housand v. Secretary of the Dep't of Health & Human Servs.*, No. 94-441V, 1996 WL 282882, at *6 (Fed. Cl. May 13, 1996) (rejecting Guillian-Barre Syndrome statistical evidence as speculative), *aff'd*, 114 F.3d 1206 (Fed. Cir. 1997).

¹⁷² See *supra* notes 22-23 and accompanying text for a discussion of intellectual due process.

and weaknesses, as does each particular study. Judges constantly must remember that automatically empowering one sort of information, or one kind of study, to the exclusion of others, makes no scientific or evidentiary sense. The rules of thumb are explained in the following sections.

1. *Statistics and Numbers Count*

In the scientific hierarchy, numbers count. Statistical analysis underlies all scientific disciplines.¹⁷³ Understanding the statistical assumptions made in extrapolating data from studies is often key to assessing the validity of a study's announced results. Scientists routinely examine the use of statistics and their underlying assumptions in critiquing each others' work, and so should judges.¹⁷⁴

Conventional wisdom dictates that large samples—studies of a large number of subjects—are better than small ones.¹⁷⁵ More trials and more studies are better than fewer.¹⁷⁶ There are several reasons for this preference for large samples and more trials. First, scientists must ensure that in collecting data from successive experiments, the researcher identified and rejected mistaken data—data that results from random¹⁷⁷ or systematic error.¹⁷⁸ Secondly, power, which is an important statistical concept,¹⁷⁹ increases with the size of the study.¹⁸⁰

¹⁷³ See Gordon J. Apple et al., *Scientific Data and Environmental Regulation*, in *Statistics and the Law* 417, 423 (Morris H. DeGroot et al. eds., 1986) (explaining importance of statistical methods for interpretation of scientific data, using example of environmental data).

¹⁷⁴ See Foster & Huber, *supra* note 37, at 148 (observing that “[s]cientists universally acknowledge the central role of statistics in their profession”); see also Colton, *supra* note 14, at 1-2 (emphasizing importance of statistical analysis to scientific argument and citing study in which 73% of published papers in three of “most frequently read medical journals” contained significant statistical errors).

¹⁷⁵ See Bailey et al., *supra* note 39, at 139 (noting, in context of epidemiology, that “researchers can increase the accuracy of the measurement of the risk of disease by enlarging the sample size”).

¹⁷⁶ See Hacking, *supra* note 71, at 149 (referring to Werner Heisenberg as solving problem of lack of knowledge in probability functions by “repeating the experiment many times”).

¹⁷⁷ Random or statistical errors are unavoidable fluctuations in the data that generally may be accounted for by statistical methods, but are otherwise beyond the control of the researcher. See Stuart L. Meyer, *Data Analysis for Scientists and Engineers* 14, 16 (1975).

¹⁷⁸ Systematic error is caused by improper experimental procedure or design and can be corrected if the cause of error is determined. Determining whether an error is random or systematic is not easy, of course, and partly for this reason proper experimental design includes running a number of tests to gather the same data. See Gad, *supra* note 81, at 5.

¹⁷⁹ See *supra* note 83 (defining power).

¹⁸⁰ See Fienberg et al., *supra* note 82, at 22 (“[T]he power of a test increases with the size of the sample . . .”). Power also increases with the degree of difference between the proposed hypothesis and the null hypothesis—the more extreme the alternatives, the better the power. See *id.*

The validity of experimental design depends on this power or sensitivity of a test to detect the response the researcher seeks.¹⁸¹ Finally, with some variability, researchers expect the data to cluster around the mean or average result in a characteristic manner.¹⁸² The standard deviation indicates how far two particular data points deviate from the mean.¹⁸³ So when the sample is small, or based on few individual measurements, the variance¹⁸⁴ and standard deviation increase, and the study may be suspect.¹⁸⁵ The reason for suspecting such a study is that small sample size decreases the power or sensitivity of the test to detect changes.¹⁸⁶ In general, a large study's findings will be more persuasive than those of a small study, and more repetition of tests is better than fewer.

Judges often think that more is better when it comes to relative risk as well. Relative risk (or the odds ratio) is a concept experts utilize in epidemiological testimony.¹⁸⁷ It is defined as the ratio of the incidence of disease occurring in people exposed to an agent to the incidence of disease in unexposed individuals.¹⁸⁸ If this ratio is greater than one, there exists a positive association between exposure and dis-

¹⁸¹ See *supra* note 83; see also Gad & Weil, *supra* note 82, at 222 ("Power for a statistical test refers to the probability that the subject hypothesis test will reject the alternative hypothesis (that there is a difference) when it is indeed false.").

¹⁸² See Gad, *supra* note 81, at 6 (describing normal distribution in which "two-thirds of all values are within one standard deviation . . . of the mean . . . and 95% are within 1.96 standard deviations of the mean"). The mean of a set of numbers "lies somewhere in the middle of the data." Kaye & Freedman, *supra* note 83, at 360.

¹⁸³ See Kaye & Freedman, *supra* note 83, at 409 (defining standard deviation as "how far a typical element deviates from the average").

¹⁸⁴ Variance is the "square of the standard deviation." *Id.* at 412.

¹⁸⁵ Another way of expressing this concept is that the confidence intervals widen. The confidence interval is an "estimate, expressed as a range, for a quantity in a population." *Id.* at 396.

¹⁸⁶ See Fienberg et al., *supra* note 82, at 22 ("[T]he power of a test increases with the size of the sample, and tests have greater power, and therefore perform better, the more extreme the alternative considered becomes."). In epidemiology, for example, very large sample sizes are necessary to detect any increased incidence of disease in the presence of exposure to a chemical agent. See Bailey et al., *supra* note 39, at 139-41 (explaining effects of sample size on accuracy); see also Philip E. Enterline, *Epidemiologic Basis for the Asbestos Standard*, *Env'tl. Health Persp.*, Oct. 1983, at 93, 96 (noting that sample of at least 1000 individuals would be necessary to detect 50% increase in deaths from asbestos).

¹⁸⁷ See, e.g., *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1315 & n.16 (11th Cir. 1999) (considering relative risk in assessing thyroid disease testimony); *Grant v. Bristol-Myers Squibb*, 97 F. Supp. 2d 986, 992 (D. Ariz. 2000) (considering relative risk in assessing and rejecting testimony on disease caused by silicone gel breast implants).

¹⁸⁸ See David E. Lilienfeld & Paul D. Stolley, *Foundations of Epidemiology* 200 (3d ed. 1994). Relative risk is thus "used to measure the strength of an association in an observational study." *Id.* For further discussion of relative risk, its calculation, and its interpretation, see *id.* at 200-02.

ease.¹⁸⁹ Judges, however, often mistakenly insist that proffered studies find a relative risk of two—a doubling of the risk—in order to be admissible.¹⁹⁰ This is mistaken because, as long as there is some increase in disease over background levels¹⁹¹ (i.e., a relative risk of greater than one, such that the incidence of the disease in the exposed population is greater than the incidence of the disease in the unexposed population), epidemiological data simply cannot predict whether exposure caused disease in any given case.¹⁹² In sum, a relative risk of greater than one is necessary before there is an epidemiological association between exposure and disease, but insistence on a particular number greater than one is unwarranted.¹⁹³

One instance where bigger is not better is statistical significance. Often referred to as the p-value, statistical significance measures the probability of getting by chance alone a test result equal to or greater than the observed result.¹⁹⁴ If the p-value of a test is five percent or less, the result is said to be statistically significant.¹⁹⁵ A p-value of one percent is considered highly significant.¹⁹⁶

Not only do numbers count, but judges should be aware that, as the old saw has it, if you torture statistics long enough, they'll tell you anything you want to hear. Common statistical errors that should set off warning signals about the validity of a study include: making inappropriate inferences based on questionable statistical techniques, failing to report contradictory evidence, and choosing statistical or graphing techniques that make the data look better than it is.¹⁹⁷ Judges, like scientists, should question whether the data was

¹⁸⁹ See Bailey et al., *supra* note 39, at 148; see also Lilienfeld & Stolley, *supra* note 188, at 200 (“The magnitude of the relative risk reflects the strength of the association; i.e., the greater the relative risk, the stronger the association.”).

¹⁹⁰ See Greenland & Robins, *supra* note 14, at 324-25 (explaining fallacy of judicial conflating of relative risk of two with probability of causation). For cases wrongly insisting upon a relative risk of two or greater, see *infra* note 206.

¹⁹¹ Background levels of a chemical are those levels to which the general population is exposed, and which cannot be attributed to any particular source. Cf. Gad, *supra* note 81, at 210 (discussing importance of concurrent controls because background tumor rates may vary over time).

¹⁹² See Greenland & Robins, *supra* note 14, at 333-34 (indicating that in certain cases, one cannot infer lack of causation from relative risk of less than two).

¹⁹³ See generally Mark Parascandola, *What Is Wrong with the Probability of Causation?*, 39 *Jurimetrics* 29 (1998) (arguing that courts are unjustified in demanding relative risk of two before finding epidemiological evidence admissible). For a general discussion of relative risk, see Harold A. Kahn & Christopher T. Sempos, *Statistical Methods in Epidemiology* 45-71 (1989).

¹⁹⁴ See Kaye & Freedman, *supra* note 83, at 402 (defining p-value).

¹⁹⁵ See *id.*

¹⁹⁶ See *id.*

¹⁹⁷ See Woolf, *supra* note 50, at 78 (discussing “deleterious effects” of research misconduct).

“trimmed” to favor a certain outcome and make sure that the results “were not the product of overenthusiastic data torture.”¹⁹⁸ Data that is “too good” is as suspect as data that does not fit the conclusions at all or is absent.¹⁹⁹ Of course, this requires judges to examine the data in the studies proffered by experts. Only by examining the data can the conclusions drawn by the studies be verified.²⁰⁰ The only way to tell if such errors are contained in an expert’s report or proffered testimony is to examine the data and the statistical inferences drawn from it, not for mathematical errors but for errors in logic.²⁰¹ Such an examination may appear a tall order for a lawyer or judge, but it is fundamental not only in assessing scientific evidence, but in analyzing information of all sorts.²⁰² Analyzing information is the heart of legal decisionmaking.

One well might question the admissibility of any statistical studies, on the grounds that they are so error-prone. Yet, these studies are a crucial part of scientific argument and can yield important insights if they are methodologically sound and based on viable assumptions.²⁰³ At a minimum, lawyers and judges involved in scientific evidence determinations need to know enough about statistics to suspect a study’s conclusions if the study included too small a number of observations

¹⁹⁸ Smith, *supra* note 166, at 1520.

¹⁹⁹ As Patricia Woolf notes, “[s]cientists are frequently suspicious of data that are ‘too good’ and thus appear to defy the ubiquitous randomness of nature; lawyers, especially in defensive postures, appear to take good data at face value.” Woolf, *supra* note 50, at 71. Moreover, she comments, “lawyers have argued that the absence of primary data strengthens the case of a scientist whose work is impugned; but scientists take missing data as a significant indicator of misconduct.” *Id.*

²⁰⁰ See Smith, *supra* note 166, at 1519 (noting that ability to “check whether the conclusions drawn from an analysis are justified increases confidence in these conclusions”).

²⁰¹ See John M. Yancey, *Ten Rules for Reading Clinical Research Reports*, 159 *Am. J. Surgery* 533, 533-38 (1990), reprinted in Foster & Huber, *supra* note 37, at 149 box 6.4.

²⁰² For example, just by reexamining the raw data of studies published by Dr. Samuel George Morton, purporting to show that black people and members of other races had smaller cranial capacity than white people, Stephen Jay Gould was able to demonstrate conclusively that Morton’s claims were groundless because the data had been misinterpreted and manipulated (consciously or unconsciously) to support his hypothesis. See Stephen Jay Gould, *The Mismeasure of Man* 68-69 (1981). Moreover, examining statistics in the scientist’s report is no more formidable than familiar legal tasks such as checking the accuracy and assumptions of accounting documents. Contradictions and inaccuracies can be revealed only by such examination.

²⁰³ See Apple et al., *supra* note 173, at 423 (providing examples of important insights statisticians can offer to legal factfinding); see also Kenneth J. Rothman, *Significance Questing*, 105 *Annals Internal Med.* 445, 445 (1986) (editorial) (indicating statistical measurement “is important because it provides the sharpest way to examine whether the deductions from theories correspond to observations”).

(or data points) or the method used to measure the data was imprecise.²⁰⁴

An emphasis on statistics does not mean that scientists can neglect the context of the system being studied, however.²⁰⁵ Epidemiological studies failing to meet the relative risk of two, an admissibility requirement for many judges, nonetheless may be extremely significant, biologically speaking.²⁰⁶ Small studies can have biological significance even without statistical significance.²⁰⁷ Both biological significance and statistical significance are important to reasoning about biological systems.²⁰⁸ It is possible to have either biological or statistical significance in the absence of the other and still reach important conclusions. Statistical significance, which is a function of the size of the study,²⁰⁹ may be present in the absence of biological significance. Conversely, biological significance may be present in the absence of statistical significance. This is because the way a given chemical works in terms of its absorption, distribution, and metabolism is as important as the statistical attributes of the study. For example, a toxicology study that reveals a rare tumor type appearing in more animals than expected, but in too few to be statistically significant, nonetheless has persuasive biological significance.²¹⁰ The results of such a study are by no means irrelevant. On the other hand, statis-

²⁰⁴ See Colton, *supra* note 14, at 142-46 (explaining how to determine correct sample size and observing that study with large standard deviation or large confidence intervals may be suspect because study sample may have been too small); Finkelstein & Levin, *supra* note 78, at 188 (noting that "large samples are needed to detect relatively small differences with adequate power").

²⁰⁵ See Gerald W. Boston, *A Mass Exposure Model of Toxic Causation: The Content of Scientific Proof and the Regulatory Experience*, 18 Colum. J. Envtl. L. 181, 255-57 (1993) (observing that scientists do not "apply the principles of significance testing in . . . an all-or-nothing dichotomizing way").

²⁰⁶ See generally Parascandola, *supra* note 193 (explaining that judges who require relative risk of two are not basing that requirement on sound science). For cases requiring a relative risk of two, see, for example, *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1315 n.16 (11th Cir. 1999) (rejecting epidemiological study that showed statistically significant correlation of silicone and disease, but had relative risk of only 1.24); *Barrow v. Bristol-Meyers Squibb Co.*, No. 96-689-CIV-ORL-19B, 1998 WL 812318, at *23 (M.D. Fla. Oct. 29, 1998) (requiring studies to show relative risk of two in order to have statistical significance, and thus be admissible, while confusing statistical significance with relative risk), *aff'd*, 190 F.3d 541 (11th Cir. 1999); *In re Hanford Nuclear Reservation Litig.*, No. CY-91-3015-AAM, 1998 WL 775340, at *8 (E.D. Wash. Aug. 21, 1998) (requiring relative risk of two for admissibility). For further discussion of relative risk, see *supra* notes 187-93 and accompanying text.

²⁰⁷ See Gad & Weil, *supra* note 82, at 222.

²⁰⁸ See *id.*

²⁰⁹ See Finkelstein & Levin, *supra* note 78, at 186-88 (observing that level of statistical significance and power increase with larger sample size).

²¹⁰ See Gad & Weil, *supra* note 82, at 222 (explaining difference between biological and statistical significance and importance of distinguishing between them, and providing examples of each).

tical significance without biological meaning can be found in clinical chemistry studies revealing biochemical parameters that are not independent of each other.²¹¹ Where damage is commonly associated with an increase in exposure to three biochemicals, for example, a statistically significant rise in just one may be biologically meaningless.²¹²

Biological and statistical significance are both important and sometimes intertwined. For example, the ability to identify an individual by using DNA has meaning only in relation to how frequently a given sequence of base-pairs can be expected.²¹³ The point is that statistical inference is meaningless without a biologically plausible theory.²¹⁴ Both provide different kinds of important information about the study.²¹⁵ Unfortunately, courts often display confusion regarding the difference between biological significance and statistical significance and the nature of different types of data.²¹⁶ The availability of a variety of studies in different disciplines showing association of agent and disease permits scientists to make valid causal inferences even in the absence of statistical significance in any particular study.²¹⁷ When faced with different types of studies, judges must categorize and assess each with an eye to both their biological and statistical significance.

2. *The Type of Study Counts*

a. Humans First. For suggesting causation of human disease, human studies are the most relevant, because studies offering the most analogous data to the hypothesis the expert espouses are the most persuasive.²¹⁸ But there are numerous kinds of human studies,

²¹¹ See, e.g., Colton, *supra* note 14, at 117 (“A result may be highly statistically significant, yet medically it may be quite unimportant.”).

²¹² See *id.*

²¹³ See *infra* note 493 (presenting basic theory of DNA identification).

²¹⁴ See Colton, *supra* note 14, at 304 (cautioning against permitting statistical analysis to generate hypotheses); Gad & Weil, *supra* note 82, at 222.

²¹⁵ See Gad & Weil, *supra* note 82, at 222.

²¹⁶ See, e.g., *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 195 (5th Cir. 1996) (rejecting testimony in absence of statistical significance); *Kelley v. American Heyer-Schulte Corp.*, 957 F. Supp. 873, 881 (W.D. Tex. 1997) (rejecting testimony based on studies failing to show statistical significance); *LeBlanc v. Merrell Dow Pharms., Inc.*, 932 F. Supp. 782, 784 (E.D. La. 1996) (granting summary judgment based on lack of “new and statistically significant studies” as required by earlier case rejecting studies proffered by plaintiff); *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 724 (Tex. 1997) (under Texas equivalent of *Daubert*, rejecting any level of statistical significance below five percent); see also *General Elec. Co. v. Joiner*, 522 U.S. 136, 145-47 (1997) (affirming trial court exclusion of expert testimony based on studies, some of which showed no statistical significance).

²¹⁷ Cf. Bailey et al., *supra* note 39, at 126 (“A strong association that is demonstrated consistently in a series of research projects leads a researcher to infer that a causal relationship exists.”).

²¹⁸ See EPA Proposed Guidelines, *supra* note 10, at 17,972-73 (“[W]hen available human data are extensive and of good quality, they are generally preferable over animal

and each of these also may be ranked according to a rule-of-thumb hierarchy. In this general hierarchy, clinical double-blind studies are best,²¹⁹ followed by cohort studies,²²⁰ with case-control studies²²¹ fairly close behind, and finally case reports.²²² Courts seem to have little difficulty accepting this ranking of human studies.²²³ Nonetheless, it is important to remember that, like any rule of thumb, this ranking is not absolute.

Another default assumption regarding human studies is that epidemiology studies on particular groups of human beings can be generalized to apply to the human population as a whole. Courts routinely

data and should be given greater weight . . ."); see also *Wade-Greaux v. Whitehall Labs. Inc.*, 874 F. Supp. 1441, 1454 (D.V.I. 1994) (recognizing hierarchy of studies in teratology), *aff'd*, 46 F.3d 1120 (3d Cir. 1994). While noting the role animal testing plays in teratological investigation, the court in *Wade-Greaux* also indicated that "human data is the glue that holds together all other data." *Id.* at 1453.

²¹⁹ The "gold standard" is the human clinical double-blind drug trial, used by the Food and Drug Administration (FDA) in its approval process, in which participants are randomly separated into groups receiving either the drug being studied or a placebo, and neither the researchers nor the participants know who is in which group (hence, "double-blind"). See Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 Nw. U. L. Rev. 643, 646-47 (1992).

²²⁰ In cohort studies,

the researcher identifies two groups of individuals: (1) individuals who have been exposed to a substance that is thought might cause a disease and (2) individuals who have not been exposed. Both groups are followed for a specified length of time, and the proportion of each group that develops the disease is compared.

Bailey et al., *supra* note 39, at 134.

²²¹ Case-control studies compare a group of individuals with a particular disease with a control group that does not have the disease. See *id.*

²²² Case reports are anecdotal filings describing "the temporal relation between an exposure and a disease." Troyen Brennan, *Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts*, 51 U. Pitt. L. Rev. 1, 46 n.192 (1989). An example is the study required by the FDA of adverse drug effects before a new drug can be marketed. Because there are few controls in case reports, courts tend to disparage these reports and give them little weight. See, e.g., *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1411 (D. Or. 1996) ("[C]ase reports and case studies are universally regarded as an insufficient scientific basis for a conclusion regarding causation because case reports lack controls." (citing cases rejecting case studies)); see also Green, *supra* note 219, at 658 (noting that problem with case reports is "that they may be purely coincidental," but observing that "case reports are sometimes validated by subsequent" studies). Although such rejection might reflect sound reasoning if case reports are the *only* evidence available, "occasionally some extremely powerful effects" may make case reports "tentatively sufficient to establish a link," particularly in conjunction with other evidence. Patricia E. Lin, *Opening the Gates to Science Evidence in Toxic Exposure Cases: Medical Monitoring and Daubert*, 17 Rev. Litig. 551, 584 (1998).

²²³ See, e.g., *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1316 (11th Cir. 1999) ("While we acknowledge the importance of anecdotal [case] studies . . . in the face of controlled, population-based epidemiological studies which find otherwise, these case studies pale in comparison.").

and without discussion employ this default assumption.²²⁴ The EPA also makes this default assumption in its risk assessments, while recognizing its limitations.²²⁵ In the absence of information to the contrary, it is a supportable default assumption for judges to assume that results of studies of a particular group showing a toxic effect from exposure to a particular agent apply to humans generally.²²⁶

The problem with human studies, however, is that they are rarely available since it is unethical to experiment on humans in all but the most limited of circumstances, and those human studies that do exist are subject to numerous imperfections.²²⁷ For example, clinical double-blind studies—the “gold standard” of human testing—are rarely available for litigation purposes.²²⁸ Human studies of any type are expensive and time-consuming.²²⁹ Moreover, even when human

²²⁴ See, e.g., *Arnold v. Dow Chem. Co.*, 32 F. Supp. 2d 584 (E.D.N.Y. 1999) (allowing admission of studies on exposed individuals without reference to applicability of study to general population); *Tyler v. Sterling Drug, Inc.*, 19 F. Supp. 2d 1239, 1241-43 (N.D. Okla. 1998) (allowing expert testimony on relationship between aspirin and Reye's Syndrome despite plaintiff having consumed aspirin for reasons different from those of humans tested in proffered human studies).

²²⁵ See EPA Proposed Guidelines, *supra* note 10, at 17,967. The assumption's limitations include, for example, the inapplicability of occupational studies focusing on healthy adult male workers to women, children, or unhealthy individuals.

²²⁶ A finding of no observed effect in human studies (as opposed to an observed effect), however, is a result that cannot be generalized necessarily. This is because a failure to disprove the null hypothesis (that there is no effect) “does not justify a conclusion that the null hypothesis has been proved.” Bailey et al., *supra* note 39, at 175; see also Enterline, *supra* note 186 (using asbestos studies as example). As one researcher explained, “[w]ith epidemiology, you can tell a little thing from a big thing. What's very hard to do is to tell a little thing from nothing at all.” Gary Taubes, *Epidemiology Faces Its Limits*, 269 *Science* 164, 164 (1995) (quoting Michael Thun). In addition to the problem of detecting small effects, there is the problem of identifying groups and individuals at high risk from exposure. See, e.g., Edward J. Calabrese, *Pollutants and High-Risk Groups* 165-66 (1978) (describing increased risk of “all people at some time” and every individual at some point). For example, epidemiology studies of worker populations would miss the effects of lead exposure on peripheral nerve function that is often observed in young children. See, e.g., Centers for Disease Control, U.S. Dep't of Health & Human Servs., *Preventing Lead Poisoning in Young Children* (1991) (discussing effects of lead studies on young children). Children are not, however, always more sensitive to chemicals than adults. For example, adults are considerably more susceptible to the renal toxicity of fluorides than are children or the elderly. See Calabrese, *supra*, at 31 (discussing effects of fluorides). Most human cancer studies were performed on adult male industrial workers who, as explained above, may not represent the exposed population. Age, sex, and individual sensitivity may all cause differences in effect. See EPA Proposed Guidelines, *supra* note 10, at 17,967 (explaining that worker populations may be sensitized to particular chemical, or that there may be sensitive subpopulations who would react differently).

²²⁷ See Goldstein & Henifin, *supra* note 123, at 187 (discussing difficulties with human studies).

²²⁸ See Green, *supra* note 219, at 647 (noting ethical difficulties associated with performing clinical double-blind studies when there is even suspicion of toxicity).

²²⁹ See Margaret A. Berger, *Eliminating General Causation: Notes Towards a New Theory of Justice and Toxic Torts*, 97 *Colum. L. Rev.* 2117, 2128 (1997) (noting that among

studies are available, they are not always the best evidence.²³⁰ For one thing, the results in human studies are nearly always equivocal because controls in human studies tend to be imperfect.²³¹ Not only are epidemiological studies notoriously poor at observing small effects in human populations, but the effects also may differ significantly on different human populations.²³² Epidemiological studies usually have low power to detect and attribute responses in a healthy population and to attribute those responses to other potentially sensitive exposed populations.²³³

In addition, each of the various kinds of human studies has its own inherent weaknesses.²³⁴ In environmental epidemiology studies, a major limitation is the lack of exposure information—with respect to both the exact chemical species to which subjects have been exposed and the actual concentrations of exposure.²³⁵ Often, complex exposures are involved, making it difficult to pinpoint a causal association.²³⁶ Thus, not only do the types of studies that judges would prefer experts to rely upon probably not exist in most cases, but those that do exist are fraught with imperfections. Therefore, judges must assess information from a variety of alternative sources in conjunction in order to decide whether there are good grounds for the experts' theory of causation.²³⁷

difficulties underlying epidemiological studies is that “considerable time and money are required to design and implement” them).

²³⁰ The Federal Judicial Center's Reference Manual on Scientific Evidence explains the potential pitfalls of human epidemiological evidence. See generally Bailey et al., *supra* note 39; see also Alvan R. Feinstein, *Scientific Standards in Epidemiologic Studies of the Menace of Daily Life*, 242 *Science* 1257, 1257 (1988) (observing that conflicting results were obtained for 56 epidemiological studies and attributing this to difficulty in conducting such studies).

²³¹ See Bailey et al., *supra* note 39, at 129 (describing important factors that cannot be controlled directly in human studies).

²³² See generally Enterline, *supra* note 186 (explaining need for relatively large increases in effect given sample sizes generally available for investigations in human studies).

²³³ See EPA Proposed Guidelines, *supra* note 10, at 17,967 (warning that generalization of human studies of particular populations might “underestimate the response of certain sensitive human subpopulations”).

²³⁴ For example, double-blind controlled experiments cannot be performed with toxic substances; cohort studies and case-control studies are subject to numerous forms of error, which are explained in numerous law review articles as well as in the Federal Judicial Center's Manual. See, e.g., Bailey et al., *supra* note 39, at 129-46; Green, *supra* note 219, at 647-53 (reviewing sources of error in human epidemiologic studies, and explaining that “[b]oth cohort and case-control studies are susceptible to a variety of errors . . . that may affect the validity of the studies' results”).

²³⁵ See Beck et al., *supra* note 149, at 23.

²³⁶ See *id.* (citing example of complex pollution indicators during 1958-1972 London fog episodes).

²³⁷ For example, in *National Bank of Commerce v. Dow Chemical Co.*, 965 F. Supp. 1490, 1507-09 (E.D. Ark. 1996), *aff'd per curiam*, 133 F.3d 1132 (8th Cir. 1998), the court was presented with a variety of studies, the cumulative impact of which it clearly did not

b. Animals and Extrapolation. In this oversimplified version of how judges should rank studies for their explanatory power respecting human effects, animal studies rank below human studies in persuasiveness. Further, primate studies are more persuasive than studies of other mammals,²³⁸ followed by bird studies, reptiles, and so on down the evolutionary or phylogenetic tree.²³⁹ In some ways, though, animal studies are better than human studies, because animal studies may have more controls and have greater sensitivity.²⁴⁰ Animal studies are also more persuasive than epidemiology in demonstrating small increases in risk.²⁴¹ To understand the mechanisms of many diseases, animal studies are imperative.²⁴² Assumptions based on phylogenetic grounds—that primates are closer phylogenetically to humans, and therefore tests done on them are the most relevant—can

understand. The court listed what it conceded were five “well-established methodologies” that could be used to assess the validity of the expert’s causation hypothesis: structure-activity relationships, *in vitro* studies (analyzing the agent’s effects on cells or tissues maintained in tissue culture), animal studies, epidemiological studies, and secular trend data. See *id.* But the court nonetheless declined to admit the plaintiff’s expert testimony because one study—the only extant epidemiological study—showed no developmental effect on the babies that had been exposed in utero. See *id.* at 1519.

²³⁸ See *Drugs and Pregnancy* 9 (Larry C. Gilstrap & Bertis B. Little eds., 2d ed. 1998) (noting that “nonhuman primates are better predictors . . . than are nonprimate models because they are phylogenetically close to humans”); Frederick A. King et al., *Primates*, 240 *Science* 1475, 1475 (1988) (explaining that “similarities in the biological mechanisms of humans and primates underlie the value of these animals for research”).

²³⁹ Cf. *Drugs and Pregnancy*, *supra* note 238, at 8-9 (discussing phylogenetic hierarchy in choice between nonhuman primate studies and rodent teratology studies); Shayne Cox Gad, *Model Selection and Scaling*, in *Animal Models in Toxicology*, *supra* note 149, at 815 (discussing basis for choice of “ideal” animal models). But see Chengelis & Gad, *supra* note 149, at 8-9 (acknowledging that “[a]ssumptions based solely on phylogenetic grounds can be quite misleading”).

²⁴⁰ See James Huff, *Chemicals and Cancer in Humans: First Evidence in Experimental Animals*, *Envtl. Health Persp.*, Apr. 1993, at 201, 204 (discussing similar metabolic processes among all mammals); see also Michael A. Kamrin, *Toxicology* 54-55 (1988) (observing that human epidemiological studies are difficult to control). By sensitivity, toxicologists mean the ability of a test to detect changes in the animal’s health. See, e.g., Gad, *supra* note 81, at 4 (defining sensitivity as “number of subjects experiencing each experimental condition divided by the variance of scores in the sample”).

²⁴¹ Even the best epidemiological cohort study may not be able to show small increases in risk. See, e.g., *Environmental Protection Agency Guidelines for Carcinogen Risk Assessment*, 51 *Fed. Reg.* 33,992, 33,995-96 (1986) (“[E]pidemiologic studies are inherently capable of detecting only comparatively large increases in the relative risk of cancer.”); N.E. Day, *Statistical Considerations*, in *International Agency for Research on Cancer, Interpretation of Negative Epidemiological Evidence for Carcinogenicity* 13, 13-14 (IARC Scientific Publications No. 65, N.J. Wald & R. Doll eds., 1985) (explaining why epidemiological data is often inconclusive).

²⁴² Cf. *Interdisciplinary Panel on Carcinogenicity*, *supra* note 152, at 686 (stating that extrapolation from animal to human “if exploited to the fullest, can provide a basis for distinguishing the degrees of risk from different carcinogens”); see also Beck et al., *supra* note 149, at 23 (explaining that each type of study offers “qualitatively different information, with unique advantages and limitations”).

be misleading.²⁴³ In some cases, the similarity between the metabolism of a particular drug in a dog and a human will be stronger than the similarity between the metabolism in a monkey and a human.²⁴⁴

Of course, animal studies have their own limitations, stemming partially from the “uncertainties in extrapolating from animals to humans” and the “uncertainties in extrapolating from the high exposures in animal studies to the lower exposures typically experienced by humans.”²⁴⁵ The point is that animal studies are not “second best.”²⁴⁶ Each type of study has its strengths and its weaknesses. Live animal studies, *in vitro* studies, and structure-activity reports—as well as human studies—all provide important information about the association of chemicals and disease.²⁴⁷ The strengths and weaknesses of the various studies complement each other, and judges are simply not scientifically justified in categorically rejecting them as evidence.

With a few notable exceptions,²⁴⁸ judges universally have difficulty understanding the persuasive impact of animal studies.²⁴⁹ Post-*Daubert* courts in toxic tort cases often find expert testimony based

²⁴³ See Chengelis & Gad, *supra* note 149, at 8-9.

²⁴⁴ See *id.* at 9 (discussing how reaction of certain chemicals in humans is more similar to reaction of those chemicals in dogs and rats than in monkeys).

²⁴⁵ Beck et al., *supra* note 149, at 25.

²⁴⁶ *Id.* (arguing that different studies provide very different kinds of information).

²⁴⁷ See *id.*

²⁴⁸ See, e.g., *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 779-80 (3d Cir. 1994). In *Paoli*, the Third Circuit held that the district court abused its discretion in excluding plaintiffs' proffered animal studies, and thus held the animal studies admissible, in part because “there is reason to think that animal studies are particularly valuable because animals react similarly to humans with respect to the chemical in question.” *Id.* at 781. Important considerations for the court were the reliance of the EPA on animal studies in the PCB context and the existence of inconclusive epidemiological data. See *id.* But see *General Elec. Co. v. Joiner*, 522 U.S. 136, 144-45 (1997) (rejecting animal study of mice exposed to PCB because “far-removed” and “dissimilar” to human exposure at issue).

²⁴⁹ See, e.g., *Raynor v. Merrell Pharms. Inc.*, 104 F.3d 1371, 1375 (D.C. Cir. 1997) (finding that animal studies have “‘testing’ problems” and noting that “[t]he only way to test whether data from non-human studies can be extrapolated to humans would be to conduct human experiments or to use epidemiological data”); *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1410 (D. Or. 1996) (finding that animal studies “are generally not considered reliable in the absence of a scientific explanation of why such extrapolation is warranted” and excluding animal studies since plaintiffs presented no such explanation); *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1480 (D.V.I.) (“The notion that one can accurately extrapolate from animal data to humans to prove causation without supportive positive epidemiologic studies is scientifically invalid because it is inconsistent with several universally accepted and tested scientific principles.”), *aff'd*, 46 F.3d 1120 (3d Cir. 1994); *Nelson v. American Sterilizer Co.*, 566 N.W.2d 671, 676 (Mich. Ct. App. 1997) (excluding expert testimony that plaintiff's liver disease was caused by low-dose exposure to chemicals merely because animal studies relied on by expert were high-dose rather than low-dose studies); *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 729-30 (Tex. 1997) (rejecting animal studies as unreliable and unable to be extrapolated to humans under Texas equivalent of *Daubert*).

primarily on animal studies inadmissible simply because the testimony relates to animals rather than to humans.²⁵⁰ This reflects a fundamental misconception about animal models, since scientists routinely generalize animal studies to human populations.²⁵¹

Animal studies are performed by exposing animals to a particular chemical and extrapolating the results to humans using what is known about the structure, function, and metabolism of the particular chemical and the similarity of the actions of the chemical in the animals studied and in human beings.²⁵² No modern scientist seriously questions the relevance of animal studies, as long as the animal studies are part of a larger story in which chemical structure, absorption, metabolism, distribution, and physiologic analogy are a part.²⁵³ Information about biological similarity and chemical structure and function (including the chemical's interaction with other chemicals, and the way the chemical acts in the body) is unquestionably important.²⁵⁴ Even

²⁵⁰ See Berger, *supra* note 229, at 2124-25 (noting courts' difficulties with admitting expert testimony based on animal studies). Even pre-*Daubert* courts using the *Frye* general acceptance standard had difficulty understanding the worth of animal studies. See, e.g., *Richardson v. Richardson Merrell, Inc.*, 857 F.2d 823, 830 (D.C. Cir. 1988) (stating that animal studies only raise "suspicion" of causation in Bendectin case); *Lynch v. Merrell-National Labs.*, 830 F.2d 1190, 1194 (1st Cir. 1987) (noting that animal studies "did not furnish a foundation for [expert's] opinions. Studies of this sort, singly or in combination, do not have the capability of proving causation in human beings in the absence of any confirmatory epidemiological data."); *In re "Agent Orange" Prod. Liab. Litig.*, 611 F. Supp. 1223, 1231 (E.D.N.Y. 1985) (commenting that data from animal studies rests on "surmise and inapposite extrapolations"), *aff'd*, 818 F.2d 187 (2d Cir. 1987).

²⁵¹ See, e.g., Chengelis & Gad, *supra* note 149, at 1 (observing that use of animal studies "in experimental medicine, pharmacological study, and toxicological assessment has become a well-established and essential practice"); Joseph M. McCune, *Animal Models of HIV-1 Disease*, 278 *Science* 2141 (1997) (noting importance of developing animal model for study of AIDS).

²⁵² See, e.g., EPA Proposed Guidelines, *supra* note 10, at 17,975-77 (discussing types and design of animal studies).

²⁵³ See Schardein, *supra* note 149, at 26-28 (observing that "not a single chemical exists that is teratogenic in the human that also has not produced malformations in *rodents*," discussing ability to extrapolate to humans from tests on other types of animals, and noting that confirming studies in multiple species increases confidence in results); see also International Agency for Research on Cancer, Preamble to 63 IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 9, 17 (1995) (noting that "it is biologically plausible and prudent" to extrapolate evidence of carcinogenicity from animals to humans and that assessment of extrapolation should consider "physical and chemical characteristics," "constituent substances," and "genetic and related effects").

²⁵⁴ See *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 747 (3d Cir. 1994) (explaining "fit" requirement that "in order for animal studies to be admissible to prove causation in humans, there must be good grounds to extrapolate from animals to humans, just as the methodology of the studies must constitute good grounds to reach conclusions about the animals themselves"). It was the lack of such an explanation that caused the dismissal of the plaintiff's case in *Joiner* and led the majority to remark in affirming the trial court's rejection of plaintiffs' studies that the studies were linked together by nothing more than the "*ipse dixit*" of the experts and that there was "simply too great an analytical gap be-

when such information is absent, however, extrapolation can provide the missing link.²⁵⁵

i. Extrapolating animal models to humans. Scientists widely assume that positive toxic effects in animal studies indicate a similar effect in humans.²⁵⁶ That does not mean that it is always true.²⁵⁷ Yet, absent information to the contrary, extrapolation is a sound assumption.²⁵⁸ It is an appropriate default because most of the data we have shows that to be the case.²⁵⁹ Therefore, judges logically should as-

tween the data and the opinion proffered." *General Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997).

²⁵⁵ For example, scientists universally accept extrapolation from rodent studies to human cancer causation as valid because "virtually all of the specific chemicals known to be carcinogenic in humans are also positive in the rodent bioassays, and sometimes even at comparable doses and with similar organ specificity." I. Bernard Weinstein, *Mitogenesis Is Only One Factor in Carcinogenesis*, 251 *Science* 387, 388 (1991). A corollary to this acceptance is that the validity of the extrapolation depends on knowing the similarities between the disease-producing mechanisms of the studied chemical in the animals studied and human beings. See generally Edward J. Calabrese, *Principles of Animal Extrapolation* (1983). Although such extrapolation is well accepted in studies of agents causing disease, extrapolating from results of animal studies of agents causing *birth defects* (called *teratogens*) to humans is more controversial. As Michael D. Green commented:

Among thirty-eight known human teratogens, thirty-seven produced an effect in at least one animal specie and twenty-seven produced an effect in more than one specie. The converse is not as accurate. Of one hundred and sixty-five agents believed not to be teratogenic in humans, forty-one percent were found to have an effect in at least one animal specie and only twenty-eight percent were negative in all species.

Green, *supra* note 219, at 655 (citing various studies). Green also observes, however, that "monkeys were a relatively good human predictor, as eighty percent of the human non-teratogens elicited a similar effect in monkeys." *Id.*; see also Gary P. Carlson, *Factors Modifying Toxicity*, in *Toxic Substances and Human Risk: Principles of Data Interpretation* 47 (Robert G. Tardiff & Joseph V. Rodricks eds., 1987) (discussing factors one must consider in extrapolating from animals to humans).

²⁵⁶ See Chengelis & Gad, *supra* note 149, at 1 ("Animals have been used as models for centuries to predict what chemicals and environmental factors would do to humans.").

²⁵⁷ For example, the utility in toxic tort litigation of animal studies designed for *regulatory risk assessment* is a frequent topic of legal controversy. See Erin K.L. Mahaney, *Assessing the Fitness of Novel Scientific Evidence in the Post-Daubert Era: Pesticide Exposure Cases as a Paradigm for Determining Admissibility*, 26 *Envtl. L.* 1161, 1182-83 (1996) (arguing for closer examination of fitness requirement and discussing ability of animal studies to meet fitness requirement).

²⁵⁸ The courts' widespread rejection of animal studies results from a failure to accept the scientifically uncontroversial default assumption that metabolic pathways in animals are similar to those in humans. See, e.g., *Nelson v. American Sterilizer Co.*, 566 N.W.2d 671, 675 (Mich. Ct. App. 1997) (noting that courts view animal studies with suspicion because "it is scientifically invalid to extrapolate observations in animal experiments directly to human beings to determine human outcomes").

²⁵⁹ Extrapolation in science is similar to a rebuttable presumption in law. That is, in the absence of contrary evidence, one is entitled to make certain assumptions—here, that animals and humans will react similarly, biologically speaking, to chemical exposure. These assumptions must be justifiable on scientific and policy grounds. The scientific grounds

sume that, in the absence of data to the contrary, animal studies can be extrapolated to humans.²⁶⁰

Despite the obvious differences between humans and laboratory animals, biochemical and metabolic processes carried out in most organs of humans and animals are similar.²⁶¹ For example, all of the chemicals recognized to cause cancer in humans also cause cancer in animals.²⁶² Chemicals that cause birth defects in humans also cause them in at least some animals.²⁶³ Studies of other kinds of diseases show a similar pattern, although they are studied less frequently.²⁶⁴ If data exist showing the opposite, then discarding the default becomes a question of the strength of the data supporting the contrary inference.²⁶⁵

Indeed, animal studies have a number of advantages. Laboratory conditions for animal studies permit the researcher to have better con-

include the overwhelming similarity of DNA, metabolism, absorption, and distribution between animals and humans. See *infra* notes 262-66 and accompanying text. An important policy justification is found in economics. If chemical manufacturers, for example, wish to rebut the extrapolation presumption, they are in a better position than litigants (or the public) to engage in research about their products. This policy would circumvent the need that a number of commentators have cited for eliminating the entire causation requirement for toxic torts. See Berger, *supra* note 229, at 2117, 2131. Courts should not reject the assumption of analogy between animals and humans for policy reasons because one of the bases of legal causation is a conclusion that "on the basis of the available evidence . . . the recurrence of that act or activity will increase the chances that the injury will also occur." Guido Calabresi, *Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr.*, 43 U. Chi. L. Rev. 69, 71 (1975) (arguing that causal links have predictive value). Probabilistically speaking, this is precisely what animal studies are designed to show, namely if the exposure event recurs, the chances of the injury recurring increase.

²⁶⁰ The arguments against using animal studies as evidence of general causation include the physiological differences between the animals studied and humans, the high dosages ordinarily used in animal studies in contrast to the low dosages typically at issue in toxic tort cases, and laboratory conditions. These are valid objections. These objections, however, are not dispositive, as high-dosage extrapolations from animals to humans can provide realistic indications of causal relationships in humans as long as the animal studies are supplemented by information regarding chemical structure; absorption, metabolism, and distribution of the chemical in the body; and the degree of physiologic analogy between the species studied. For an outline of factors of extrapolation, see Carlson, *supra* note 255, at 47-70.

²⁶¹ See Cranor et al., *supra* note 39, at 50 (stating that there are more physiologic, biochemical, and metabolic similarities between laboratory animals and humans than there are differences).

²⁶² See International Agency for Research on Cancer, *supra* note 253, at 17.

²⁶³ See *supra* notes 253-55 (noting strength of extrapolation but acknowledging its limitations).

²⁶⁴ See, e.g., Goldstein & Henifin, *supra* note 123, at 191 ("If a heavy metal such as mercury causes kidney toxicity in laboratory animals, it will almost certainly do so at some dose in humans.").

²⁶⁵ See Popper, *supra* note 36, at 399 (explaining interrelationship of empirical data to informative content); see also EPA Proposed Guidelines, *supra* note 10, at 17,965 (discussing circumstances under which default assumptions should be employed).

trols over the experimental conditions.²⁶⁶ The higher dosages normally employed in animal studies more accurately reveal the presence of a dose-response relationship. Also, the time required to conduct an animal study is much shorter than for human studies.²⁶⁷ Thus, animal studies are often the primary source of information regarding the impact of chemicals on humans.²⁶⁸ Differences between animals and humans in metabolism, body size, dose, life span, and other factors should be accounted for, but courts should not reject animal studies categorically as they now often do.²⁶⁹

ii. *The quantitative extrapolation: extrapolating high doses in animals to low doses in humans.* There are two fundamental types of extrapolation from animals to humans.²⁷⁰ The *qualitative* extrapolation discussed in the previous section—if a chemical causes an effect in one species, then it also will cause it in humans—is well-accepted by scientists. The *quantitative* extrapolation, represented by the dose-response curve, addresses the specific level at which the chemical will have an effect.²⁷¹ There is more debate over this quantitative estimation of the magnitude of dose producing a similar effect in animals and humans.²⁷² Nonetheless, this debate is generally resolved in favor

²⁶⁶ See Brennan, *supra* note 68, at 507 (noting that, as opposed to animal studies, “[r]esearchers cannot control the factors that affect the quality of the data” in epidemiologic studies).

²⁶⁷ See Green, *supra* note 219, at 654 (“[M]any animal species reproduce readily and have short life cycles, thereby reducing the time required to conduct an experimental study.”).

²⁶⁸ See Risk Assessment in the Federal Government, *supra* note 168, at 22 (“The inference that results from animal experiments are applicable to humans is fundamental to toxicologic research Notwithstanding uncertainties associated with interpretation of some animal tests, they have, in general, proved to be reliable indicators of carcinogenic properties”).

²⁶⁹ But see Jack L. Landau & W. Hugh O’Riordan, *Of Mice and Men: The Admissibility of Animal Studies to Prove Causation in Toxic Tort Litigation*, 25 *Idaho L. Rev.* 521, 565 (1998-99) (“Animal studies have no place in the courtroom.”). Indeed, for most toxic tort cases, there is a paucity of data, not necessarily because there is no causation, but because “many toxic tort cases share . . . a history of manufacturer neglect in conducting basic safety research on the safety of their products.” Rebecca S. Dresser et al., *Breast Implants Revisited: Beyond Science on Trial*, 1997 *Wis. L. Rev.* 705, 732-33 (noting that asbestos, tobacco, Agent Orange, Dalkon Shield, ultra-absorbent tampon, and Bendectin litigations all suffered from inadequate manufacturer testing).

²⁷⁰ See Interdisciplinary Panel on Carcinogenicity, *supra* note 152, at 685 (discussing two types of extrapolation).

²⁷¹ See Goldstein & Henifin, *supra* note 123, at 191.

²⁷² See Michael A. Dorato & Mary Jo Vodicnik, *The Toxicological Assessment of Pharmaceutical and Biotechnology Products*, in *Principles and Methods*, *supra* note 82, at 189, 193 (explaining that qualitative extrapolation is more reliable than quantitative extrapolation); David G. Hoel et al., *Implication of Nonlinear Kinetics on Risk Estimation in Carcinogenesis*, 219 *Science* 1032, 1032 (1983) (discussing extrapolation controversy and suggesting that studies of chemical metabolism and pharmacokinetics may be helpful); In-

of quantitative extrapolation, because animal testing of chemical effects (including quantitative extrapolation) predicts human toxicity in all but ten percent of comparisons.²⁷³

In other words, the corollary assumption to the applicability of animal studies to human populations, that high dosage studies can be extrapolated to the low dosages more commonly found in human environmental exposure, is well-substantiated.²⁷⁴ Scientists understand that high dosages for short time periods are roughly equivalent to low dosages over extended periods.²⁷⁵ In addition, it is known that "as animal species become larger, they also become more sensitive to short term toxicities."²⁷⁶ Humans can be many times more sensitive than experimental animals.²⁷⁷ All these factors indicate that judges should adopt as a rebuttable presumption that, in the absence of data to the contrary, high dosage studies validly can be extrapolated to the low dosages found in most environmental exposures.²⁷⁸

terdisciplinary Panel on Carcinogenicity, *supra* note 152, at 685-86 (discussing uncertainties in quantitative extrapolation).

²⁷³ See Dorato & Vodcnik, *supra* note 272, at 193; see also Interdisciplinary Panel on Carcinogenicity, *supra* note 152, at 685-86 (noting uncertainties of extrapolation from high to low doses, but concluding that reliable estimates of risk are nonetheless possible, especially if combined with information about metabolism and pharmacokinetics).

²⁷⁴ See EPA Proposed Guidelines, *supra* note 10, at 17,967 (stating that default assumption accepted by many scientists is that "effects seen at the highest dose tested are appropriate for assessment" while acknowledging that "it is necessary that the experimental conditions be scrutinized" when making such assumption).

²⁷⁵ See, e.g., J. Carl Barrett, Mechanisms of Multistep Carcinogenesis and Carcinogen Risk Assessment, *Env'tl. Health Persp.*, Apr. 1993, at 9, 9-20; Hoel et al., *supra* note 272, at 1032 (explaining that standard method of testing of chemical effects of high doses—which necessitates extrapolation to low doses common for environmental exposure—is necessary because of impracticably large number of animals that would be required for low-level testing and suggesting that pharmacokinetics may add valuable information to extrapolation); R.L. Melnik et al., Cell Proliferation and Chemical Carcinogenesis: A Symposium Overview, 7 *Molecular Carcinogenesis* 135, 135-38 (1993). But see Philip H. Abelson, Risk Assessments of Low Level Exposures, 265 *Science* 1507, 1507 (1994) (editorial) (urging development of low-level tests because linear extrapolation implies that "no dose, however small, is safe" and because "[e]xamples of instances in which these assumptions are invalid are becoming numerous").

²⁷⁶ Gad, *supra* note 239, at 826 (attributing higher sensitivity in larger species to "increases of available target tissues and decreases in metabolic rate as size increases").

²⁷⁷ See Dorato & Vodcnik, *supra* note 272, at 193 ("Humans can be as much as 50 times more sensitive on a milligram/kilogram (mg/kg) basis than experimental animals."); see also Gad, *supra* note 239, at 826 (observing that some scientists "believe that humans are more sensitive than any test species, even if that species is larger than humans").

²⁷⁸ The overwhelming consensus on correlating an observed dose response curve to a relationship at lower doses is to use a linear model. See EPA Proposed Guidelines, *supra* note 10, at 17,968 (adopting linear default approach in absence of data); Hoel et al., *supra* note 272, at 1032 (noting that "[t]he standard way of doing the low-dose extrapolation has been to assume a parametric model relating applied dose to tumor response, and to use the high-dose data to estimate the model parameters"). This view should not be accepted merely because it is prevalent, of course, but because the theory is falsifiable, has been

Biological damage processes (as we currently understand them) fall into two broad classes: those implying either threshold or non-threshold responses.²⁷⁹ For systemic toxicants (as opposed to carcinogens), the threshold response theory asserts that “there is a level of exposure below which there is minimal, if any, chance for an adverse effect.”²⁸⁰ Under this hypothesis, “multiple cells must be injured before an adverse effect is experienced and . . . the injury must occur at a rate that exceeds the rate of repair.”²⁸¹ As long as the dose is small enough, the threshold theory says that adaptive processes will repair any damage.²⁸² Courts are familiar with this theory, and frequently use it to reject expert testimony where the plaintiff’s precise exposure level is unknown, citing Paracelsus’s maxim that the dose determines the poison.²⁸³

Nonthreshold effects, on the other hand, are associated with diseases like cancer that are caused by genetic mutations, and do not

subjected to testing, and remains the best explanation for the observed data. Cf. Popper, *supra* note 36, at 121 (“[T]hose theories should be given preference which can be most severely tested . . .”).

²⁷⁹ See D. Cooper Rees & Dale Hattis, *Developing Quantitative Strategies for Animal to Human Extrapolation*, in *Principles and Methods*, *supra* note 82, at 275, 276.

²⁸⁰ Beck et al., *supra* note 149, at 40. The prevalent theory for the action of carcinogens, on the other hand, is that they are nonthreshold. See *infra* notes 284-88 and accompanying text.

²⁸¹ Beck et al., *supra* note 149, at 40.

²⁸² See Rees & Hattis, *supra* note 279, at 276 (explaining that as long as dose of toxin is “small enough that it does not disturb the homeostatic relationships beyond specific limits, it is assumed that adaptive processes will repair any damage that may have been temporarily produced”).

²⁸³ Paracelsus was a Renaissance alchemist who believed in the doctrine of signatures. This doctrine demanded that diseases be treated with remedies similar in shape to the affected organ. For example, to cure a liver ailment, he would propose to treat it with an herb shaped like a liver. See Hacking, *supra* note 71, at 42. Despite the long-ago debunking of the doctrine of signatures, courts continue to cite Paracelsus as authority. See, e.g., *National Bank of Commerce v. Dow Chem. Corp.*, 965 F. Supp. 1490, 1506 (E.D. Ark. 1996) (relying on Paracelsus for notion that “toxicity is a function of dose” (quoting *Cavallo v. Star Enter.*, 892 F. Supp. 756, 769 n.27 (E.D. Va. 1995), *aff’d in part and rev’d in part* on other grounds, 100 F.3d 1150, 1159 (4th Cir. 1996), cert. denied, 522 U.S. 1044 (1998))), *aff’d per curiam*, 133 F.3d 1132 (8th Cir. 1998); *Cartwright v. Home Depot U.S.A., Inc.*, 936 F. Supp. 900, 906 (M.D. Fla. 1996) (same); *Cavallo*, 892 F. Supp. at 769 n.27 (same); *Carroll v. Litton Sys., Inc.*, No. B-C-88-253, 1990 WL 312969, at *74 & *96 n.7 (W.D.N.C. Oct. 29, 1990) (relying on Paracelsus for idea that “all substances are poisons and the right dose differentiates between a remedy and a poison”), *aff’d in part and rev’d in part mem.*, 47 F.3d 1164 (4th Cir. 1995). Why modern courts cite his maxim without understanding his theory is a perplexing question. The classic toxicology argument states that “most chemicals exhibit a dose-response relationship, resulting in an apparent or effective threshold for at least some agents.” Gad, *supra* note 81, at 235 (presenting common arguments). Yet this is untrue with respect to at least some carcinogens. See *id.* at 234-38 (summarizing arguments for and against threshold responses).

depend on the size of the dosage.²⁸⁴ The nonthreshold, or linear effects, theory holds that “even a single molecule of a DNA-reactive substance has a small but nonzero chance of setting in motion a chain of events leading to mutagenesis or carcinogenesis.”²⁸⁵ In addition, the threshold dose notion is scientifically questionable as applied to carcinogens because, among other reasons, it is impossible to demonstrate a no-effects level mathematically.²⁸⁶ Further, “there is . . . no empirical or theoretical basis for determining the dose at which [this threshold] may occur.”²⁸⁷ Although there is some evidence that not all cancers are mediated by direct action on the genes, and therefore some carcinogens may have threshold effects, for most diseases, including cancer, a nonthreshold theory is the most plausible default in the absence of specific data to the contrary.²⁸⁸

Unfortunately, even if we find the threshold theory of biological responses persuasive, there is usually not enough information to demonstrate either threshold or nonthreshold responses (both for cancer and other diseases).²⁸⁹ Because we rarely have biologically-based or case-specific models for the levels both above and below the threshold, we typically must assume that all diseases operate on a nonthreshold basis (the linear default approach).²⁹⁰ Although the re-

²⁸⁴ See Gad, *supra* note 81, at 234-35 (summarizing arguments against existence of threshold for carcinogens and opining that arguments for threshold are “more mechanistic”); Rees & Hattis, *supra* note 279, at 276 (noting that threshold processes are “generally assumed to underlie gene mutation, and, by extension, that portion of carcinogenesis that is produced by gene and chromosomal mutations”).

²⁸⁵ Rees & Hattis, *supra* note 279, at 277. Mutagenesis is the process of genetic mutations, while carcinogenesis is the process of cancer formation.

²⁸⁶ See R.W. Hart & L. Fishbein, *Interspecies Extrapolation of Drug and Genetic Toxicity Data*, in 1 *Toxicological Risk Assessment* 3, 32 tbl.19 (D.B. Clayson et al. eds., 1985) (presenting arguments against “threshold dose” theory).

²⁸⁷ *Id.* Gad points out that the huge numbers of animals that would be required for testing of low-level doses in order to test the threshold theory makes such experimentation infeasible. See Gad, *supra* note 81, at 233.

²⁸⁸ For arguments in favor of and against “threshold dose” theory, see Hart & Fishbein, *supra* note 286, at 31 tbl.18, 32 tbl.19. It is important to note that even if there were a threshold response, that “would actually only mean that a linear (or any other dose-response relationship) would start at some point above zero, being discontinuous only in the extreme lowest dose range.” Gad, *supra* note 81, at 234. Moreover,

the presence of a background exposure of carcinogens and promoters, and of spontaneously occurring cancers in a population at risk as large and diverse as that of human beings, implies that even if there are thresholds for some or most individuals, there will still remain others who have been “jumped over” their individual thresholds by background events.

Id.

²⁸⁹ See Hart & Fishbein, *supra* note 286, at 32 tbl.19.

²⁹⁰ See *id.* In Popperian terms, the competing linearity and threshold theories need to be continually reevaluated in light of the severity of criticism each has withstood (or can withstand). See Popper, *supra* note 96, at 368; Popper, *supra* note 36, at 170 n.2, 418. However, until more studies are performed that support the threshold theory, the linearity

ceived wisdom is that linear dose-response assumptions are conservative in that they overestimate risks to protect the public health, there is evidence that they actually may underestimate risk at low doses.²⁹¹ Moreover, as a policy matter, a linear presumption in toxic tort cases is fairer to plaintiffs, because information about exposure levels is rarely available.²⁹²

Nonlinear approaches therefore should be limited to those situations in which adequate data on the mode of action (that is, how the chemical affects bodily processes and the progression of disease) show that linearity is not reasonable.²⁹³ Thus, in the absence of information to the contrary, a nonthreshold response is generally assumed by scientists and therefore should be assumed by judges.²⁹⁴

iii. Extrapolating to different target organs. Not only must the results of high dosage animal studies be extrapolated to the low dos-

model (which has more empirical support, with a few documented exceptions) is a solid working theory. See, e.g., Lakatos, *supra* note 58, at 14 (explaining that Newtonian physics worked well as theory, despite documented anomalies, until Einstein's explanation supplanted it). The EPA Proposed Guidelines suggest, in the absence of data supporting a case-specific or biologically-based model, that a curve-fitting model be used for the observed range of data and a linear model for extrapolating to the unobserved range. See EPA Proposed Guidelines, *supra* note 10, at 17,968. The EPA Proposed Guidelines explain that "the linear approach is to draw a straight line between a point of departure from the observed data . . . and the origin (zero dose, zero response)." *Id.*

²⁹¹ See Leslie Roberts, *Is Risk Assessment Conservative?*, 243 *Science* 1553, 1553 (1989) (citing studies on vinyl chloride).

²⁹² For arguments in favor of placing the burden on defendant manufacturers to offer contrary information to defeat the presumption, see Beecher-Monas, *supra* note 8, at 1039-92; Berger, *supra* note 229, at 2140-52; Wendy E. Wagner, *Choosing Ignorance in the Manufacture of Toxic Products*, 82 *Cornell L. Rev.* 796, 810-32 (1997). For cases excluding expert testimony where plaintiffs did not know their exposure level, see, for example, *Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 810 (6th Cir. 1994) (affirming exclusion of expert testimony that chlordane, carcinogen acknowledged to be applied in manner that violated federal criminal law, caused plaintiffs' injuries where testimony was based on animal studies that used high-dosage levels rather than low-dosage levels to which plaintiffs were exposed); *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490 app. F at 1552-54 (E.D. Ark. 1996) (discrediting plaintiff's expert testimony because defendants' expert proffered single epidemiology study showing no effects, even though plaintiff offered evidence to show study's results were inconclusive because it involved only five subjects who were exposed to unknown pesticides at unknown levels), *aff'd per curiam*, 133 F.3d 1132 (8th Cir. 1998); *Cavallo v. Star Enter.*, 892 F. Supp. 756, 766 (E.D. Va. 1995) (finding testimony inadmissible because expert could not identify plaintiff's exposure levels or dose-response relationship with respect to jet fuel at issue), *aff'd in part and rev'd in part on other grounds*, 100 F.3d 1150, 1159 (4th Cir. 1996), *cert. denied*, 522 U.S. 1044 (1998); *Schmaltz v. Norfolk & W. Ry. Co.*, 878 F. Supp. 1119, 1122 (N.D. Ill. 1995) (excluding testimony where expert did not know level of plaintiff's exposure to herbicide); *Chikovsky v. Ortho Pharm. Corp.*, 832 F. Supp. 341, 345-46 (S.D. Fla. 1993) (excluding testimony where expert did not know how much Retin-A was absorbed by plaintiff or dosage level at which Retin-A becomes unsafe for pregnant women).

²⁹³ See *supra* note 286 and accompanying text.

²⁹⁴ See *id.*

ages common to human exposure, but one also must account for the possibility that a different organ might be affected in the animal studies than was affected in the plaintiff. Whether the effects on the organ in the studies (the target organ) may be extrapolated to a different organ in human beings is controversial. Most scientific knowledge indicates that target organs should be the same in the animals studied and in humans.²⁹⁵ There is typically a close correspondence between the target organ in human studies and the organ in at least one of the animal species studied.²⁹⁶ Still, various tissues react differently across species.²⁹⁷ Thus, there might be justifications in a particular case for scientists and judges to adopt either the assumption that the affected organs must be the same, or the contrary assumption that they may be different.²⁹⁸ The EPA has taken the default position that, with respect to cancer, in the absence of contrary data it does not matter which organ is affected; if an animal develops cancer in any organ, the agent will be assumed to be generally carcinogenic in humans.²⁹⁹ Under this assumption, demonstrating the development of lung cancer in mice, for example, would be admissible testimony where human cancer in any organ was at issue. Because we know so little about cancer causation, there is justification for this assumption at least with respect to cancer.

On the other hand, the basis for the EPA's position that a carcinogen in an animal organ can be a carcinogen in any human organ is that this assumption protects the public health and minimizes false negatives.³⁰⁰ Judicial decisionmakers must strive to minimize both false negatives *and* false positives, so the EPA's assumption should

²⁹⁵ See EPA Proposed Guidelines, *supra* note 10, at 17,967 (stating that “[t]arget organs of carcinogenesis for agents that cause cancer in both animals and humans are most often concordant at one or more sites” across species).

²⁹⁶ See *id.*; see also Huff, *supra* note 240, at 204 (discussing similar metabolic processes among all mammals and noting that “mechanisms of chemical toxicity are, to a large extent, identical in animals and man”).

²⁹⁷ See Huff, *supra* note 240, at 204 (comparing similarities in human and animal cancer development).

²⁹⁸ The Interdisciplinary Panel on Carcinogenicity explains the basis for assuming that target organs need not be identical: “A malignant neoplasm [cancer] . . . displays, or it has the capacity for, invasion of normal tissues, metastasis, and causing death to the host.” Interdisciplinary Panel on Carcinogenicity, *supra* note 152, at 686 n.4 (internal quotation marks omitted). This capacity for invading normal tissues is what accounts for the possibility of different target organs in cancer. Because “benign neoplasms may represent a stage in the evolution of a malignant neoplasm,” finding tumors in any organ may be sufficient justification to assume the affected organs may be different. *Id.* (internal quotation marks omitted).

²⁹⁹ See EPA Proposed Guidelines, *supra* note 10, at 17,967 (“The default assumption is that positive effects in animal cancer studies indicate that the agent under study can have carcinogenic potential in humans.”).

³⁰⁰ See *id.*

not be adopted wholly by judges. As a policy matter, a more justifiable assumption for the legal system is that, absent evidence showing a range of target organs, the target organ affected in animal studies should be the same as that affected in the injured plaintiff.³⁰¹ A position that the organs must be the same is more defensible (both scientifically and for legal purposes) because the metabolic route to those organs usually will be similar in animals and humans. Therefore, if the expert relies on animal studies showing an effect on different organs than those affected in the plaintiff, the studies would be inadmissible in the absence of further explanation.

iv. Route of exposure extrapolation. A further extrapolation involves deciding whether exposure to a chemical in one manner (e.g., through the skin) is equivalent to exposure in a different manner (e.g., by ingestion). We call this the route of exposure extrapolation. The three main exposure routes in humans, either alone or in combination, are ingestion, inhalation, and dermal (skin) absorption. The route of administration of a chemical in laboratory animal experiments may be different from the route of exposure of the plaintiff: The agent may have been *injected* into the animals, for example, whereas the human plaintiff's *skin* was exposed to the chemical.³⁰²

No matter what the route of exposure, however, the critical measurement in a toxicological study is the bloodstream concentration of the chemical.³⁰³ The level of chemical in the target organ is generally

³⁰¹ That is, just because in most cases metabolic processes are similar, animal studies showing, for example, an agent's toxic effects on the lungs should not be assumed by judges to be relevant to the issue of whether a plaintiff's *pancreatic* disease (other than cancer) was caused by the agent. The expert must provide other studies showing a different metabolic pathway in humans that would make it likely that the agent would affect both the lungs in mice and the pancreas in humans. The EPA Proposed Guidelines recognize that the default of extrapolation to different target organs adopted by the Guidelines may not be appropriate when data allows the consideration of "the influences of route of exposure, metabolism, and particularly, hormonal modes of action." EPA Proposed Guidelines, *supra* note 10, at 17,968. Therefore, a more appropriate assumption for the legal system is that target organs should be similar.

³⁰² See, e.g., *General Elec. Co. v. Joiner*, 522 U.S. 136, 144 (1997) (approving district court's rejection of studies of mice that developed cancer after being *injected* with PCBs to support plaintiff's assertion that PCB exposure to plaintiff's *skin* caused cancer).

³⁰³ See Andrew Gordan Renwick, *Toxicokinetics—Pharmacokinetics in Toxicology, in Principles and Methods*, *supra* note 82, at 106-07 (explaining that slope of plasma-concentration time-curve is key measurement in toxicological studies). Moreover, if "dosage is adjusted so that equivalent serum concentrations are maintained, [interspecies] differences in response tend to disappear." Steven E. Mayer et al., *Introduction: The Dynamics of Drug Absorption, Distribution, and Elimination*, in Goodman & Gilman's *The Pharmacological Basis of Therapeutics* 1, 26 (Alfred Goodman Gilman et al. eds., 6th ed. 1980). Thus, absorption is equivalent to "the appearance of the substance in the circulation," and the "rate of absorption can be determined from the plasma concentration time-curve" of the chemical in question. World Health Org., *Principles of Toxicokinetic Studies* 28 (Env'tl.

a function of its concentration in the blood.³⁰⁴ Aside from local effects at the site of contact,³⁰⁵ a toxic substance can cause injury only after absorption by the body.³⁰⁶ Absorption, regardless of the site of contact, is dependent upon solubility of the chemical in question.³⁰⁷ Thus, for example, asbestos (which is relatively insoluble) has a greater effect when inhaled than when ingested because the particles stay in the lungs longer (and therefore dissolve more) than they do in the gut.³⁰⁸ But if one accounts for solubility differences, the route of exposure makes little difference.³⁰⁹ Extrapolation from one exposure route to another can be accounted for mathematically through physiologically based pharmacokinetic models,³¹⁰ even though such models

Health Criteria 57, 1986). In sum, the limiting factor for chemical effects is that the chemical "must be present in appropriate concentrations at its sites of action," Leslie Z. Benet et al., *Pharmacokinetics: The Dynamics of Drug Absorption, Distribution, and Elimination*, in Goodman & Gilman's *The Pharmacological Basis of Therapeutics* 3, 3 (Alfred Goodman Gilman et al. eds., 9th ed. 1996) [hereinafter Goodman & Gilman's (9th ed.)], and this depends principally on the amount of chemical that gets into the bloodstream, see id. fig.1-1.

³⁰⁴ See Frank C. Lu, *Basic Toxicology: Fundamentals, Target Organs, and Risk Assessment* 23 (3d ed. 1996).

³⁰⁵ A local effect is one that occurs at the site of first contact with a chemical. See Curtis D. Klaassen, *Principles of Toxicology*, in Goodman & Gilman's (9th ed.), supra note 303, at 63, 66. An example of a local effect is the injury to the skin caused by exposure to tetramethyl lead. See id. A systemic effect is, on the other hand, one that occurs after absorption and distribution of the chemical. See id. The two categories "are not mutually exclusive." Id. For example, tetramethyl lead causes both local effects at the site of contact (skin) and systemic effects on the central nervous system. See id. Most toxic substances produce systemic effects. See id. When substances have a predominantly local effect, "the frequency of tissue reaction depends largely on the portal entry (skin, gastrointestinal tract, or respiratory tract)." Id. Because most toxic tort cases involve systemic rather than local effects, the assumption that the route of exposure is of minimal importance is a sound one. If, on the other hand, the defense can demonstrate either predominantly local effects or that the systemic effects of the chemical in question are route-dependent, then the default presumption would be rebutted.

³⁰⁶ See Lu, supra note 304, at 13.

³⁰⁷ See Mayer et al., supra note 303, at 5 (discussing factors that modify absorption of drugs).

³⁰⁸ See William E. Pepek, *Effect of Exposure Route on Potency of Carcinogens*, 13 *Reg. Toxicology & Pharmacology* 3, 11 (1991) (discussing potency of carcinogens inhaled versus carcinogens ingested).

³⁰⁹ See id. at 13.

³¹⁰ Pharmacokinetics refers to the way a chemical is absorbed, transported through the bloodstream to the target organ, metabolized, and excreted. See generally Charles Mann, *Women's Health Research Blossoms*, 269 *Science* 766, 768 (1995) (discussing pharmacokinetics in relation to gender); Mayer et al., supra note 303 (discussing pharmacokinetics); Renwick, supra note 303, at 101 (defining pharmacokinetics). Pharmacokinetic models "describe . . . processes affecting the disposition of a chemical and its metabolism from the time it is absorbed to its interaction with different and various body tissues." Beck et al., supra note 149, at 39. These models attempt to account for "variation[s] with time of drug concentration in the blood, serum, or plasma as a result of absorption, distribution, and elimination." Mayer et al., supra note 303, at 21 (emphasis omitted) (describing elementary kinetic model of fundamental pharmacokinetics); see also

themselves use many assumptions and simplifications and are more accurate when data is available with respect to exposure level and toxic effect.³¹¹ In the absence of such information, therefore, a fair assumption is that exposure routes are equivalent.

In assessing differences in the way the chemical agent was applied in the animal studies and in exposed humans, the EPA uses the default assumption that, for carcinogens, exposure by one route is equivalent to exposure by another route.³¹² The EPA recognizes that as long as the internal dose is the same, there is no reason to suspect the chemical agent will act differently if it was absorbed into the body through the skin or taken internally.³¹³ The metabolism of the internal dose will be the same.³¹⁴ In other words, “[m]easurement of plasma concentrations provides much needed proof of absorption and exposure.”³¹⁵

The assumption that chemical agents act the same, regardless of application, not only protects the public health to some degree, but it also offers the soundest scientific explanation of how things work.³¹⁶ Absorption and distribution are similar in all mammalian species.³¹⁷ Toxicologists note “that almost every toxicant can pass through one or more portals of entry, although there may be considerable differences in rate.”³¹⁸ Unquestionably, data is needed with respect to absorption

Renwick, *supra* note 303, at 106-07 (discussing mathematical model to describe changes in blood plasma levels).

³¹¹ See Kannan Krishnan & Melvin E. Anderson, *Physiologically Based Pharmacokinetic Modeling in Toxicology*, in *Principles and Methods*, *supra* note 82, at 149, 174 (arguing that “appropriate equations to represent each exposure pathway” may account adequately for exposure route differences and allow extrapolation).

³¹² See EPA Proposed Guidelines, *supra* note 10, at 17,978 (adopting default assumption that exposure routes are equivalent, and stating that “if it is not contrary to available data, it is assumed as a default that toxicokinetic and metabolic processes are qualitatively comparable between species”).

³¹³ See *id.*

³¹⁴ That is, “[t]o produce its characteristic effects, a [chemical] must be present in appropriate concentrations at its sites of action” within the cell. Benet et al., *supra* note 303, at 3. See generally Rees & Hattis, *supra* note 279 (explaining basic principles of pharmacokinetics). This is particularly true for carcinogenic chemicals, where dose route has little effect. See Pepelko, *supra* note 308, at 11.

³¹⁵ Dorato & Vodcnik, *supra* note 272, at 212.

³¹⁶ See Chengelis & Gad, *supra* note 149, at 9 (noting that route of exposure makes little difference and explaining that important factor in toxicology is access to circulatory system); W.E. Pepelko, *Feasibility of Route Extrapolation in Risk Assessment*, 44 *Brit. J. Indus. Med.* 649, 649 (1987) (explaining necessity of route extrapolation and use of uncertainty factors when environmental exposure differs from experimental route of exposure).

³¹⁷ See Chengelis & Gad, *supra* note 149, at 8 (“The principles that govern absorption and distribution apply fairly equally across all species.”). There are metabolic rate differences, but most of these can be accounted for mathematically. See *id.*

³¹⁸ Ernest Hodgson & Patricia E. Levi, *Absorption and Distribution of Toxicants*, in *A Textbook of Modern Toxicology*, *supra* note 149, at 27 (explaining that “few, if any, chemicals will be excluded from entry” to body because of route of exposure).

characteristics and toxicity across routes.³¹⁹ The “[d]evelopment of scientifically based principles, procedures, and data for route-to-route [exposure] extrapolations” would undoubtedly improve their validity.³²⁰ But in the absence of such information, the default assumption of equivalence is justified as simply the best we can do.

In addition, this judicial assumption makes sense from a policy standpoint since it encourages manufacturers to engage in safety research.³²¹ If manufacturers wish to demonstrate that there is no equivalence in exposure routes, they are in the best position to produce the data to support that counterassumption.³²² They are in a better position to produce the necessary data than are consumers since they are in the business of producing and testing the chemical.³²³ From a policy perspective, placing the burden on plaintiffs of proving in each case—without access to research facilities—that exposure by different routes would be the same makes little sense.³²⁴

³¹⁹ See Summary Report of the Workshops on Principles of Route-to-Route Extrapolation for Risk Assessment, in *Principles of Route-to-Route Extrapolation for Risk Assessment* 1, 2-3 (Timothy R. Gerrity & Carol J. Henry eds., 1990) [hereinafter Summary Report] (“At present little is known about absorption characteristics, the potential for portal-of-entry effects, and the potential for first-pass metabolic effects for most compounds by most routes.”).

³²⁰ *Id.* at 3.

³²¹ Pharmacokinetic information on most chemicals is simply not available. See William E. Peipelko & James R. Withey, *Methods for Route-to-Route Extrapolation of Dose*, 1 *Toxicology & Indus. Health* 153, 159 (1985) (discussing lack of data to compare “toxic effects of similar doses administered by different routes”). Without manufacturer research, no data on route equivalence may be gathered at all. Thus, the use of the exposure route equivalence assumption is an efficient allocation of resources, placing the cost on the party best able to assume it. See Wagner, *supra* note 292, at 797-803 (arguing for burden on manufacturer to do safety research).

³²² Most chemicals on the market never have been tested for safety. See Beecher-Monas, *supra* note 8, at 1090 & nn.261-63. It does not seem unfair, therefore, once safety has become an issue, to place the burden of producing evidence on those who profit from the manufacture of the chemicals, at least to the extent of refuting a presumption of route equivalence. This is a particular concern where little is known about the pharmacokinetics of particular chemicals and yet those compounds are marketed widely. See Summary Report, *supra* note 319, at 3 (discussing lack of knowledge about route-to-route extrapolation).

³²³ Thus the burden of production of evidence to refute the presumption of exposure-route equivalence should be placed on the defendant chemical marketer because of the inequality of resources between plaintiff and defense. See Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 *Stan. L. Rev.* 1477, 1543 (1999) (arguing that burdens of persuasion and production are economizing devices and therefore should be placed on party with greatest access to resources).

³²⁴ See Peter C. Carstensen, *Explaining Tort Law: The Economic Theory of Landes and Posner*, 86 *Mich. L. Rev.* 1161, 1181 (1988) (noting that “an efficiency-increasing solution is to identify a class of participant which has distinct cost advantages in handling any aspect of the accident problem”).

c. *In Vitro Studies*. Less persuasive, as a general matter, than (in vivo) animal studies are in vitro tissue culture studies designed to examine the effects of agents on isolated cells, bacteria, organs, or embryos.³²⁵ While they may be ranked lower in this rule-of-thumb hierarchy, these tests can help to determine the effect of a chemical at the cellular level.³²⁶ In vitro studies are used to support the interpretation of information from human and animal studies.³²⁷ Bearing in mind that the “rational approach is to examine all sources of information in the evaluation of toxic chemicals,”³²⁸ in vitro studies are not “inferior” studies, for they are studies that add vital information to the mix. Although they may be insufficient standing alone, these studies may explain mechanisms of action that other types of studies cannot. Courts, however, are even less comfortable when experts rely on these studies than they are with in vivo animal studies.³²⁹ It is a mistake to disregard these studies because these studies are invaluable in understanding metabolic processes at the cellular level. If such studies have been performed, they add explanatory power to the expert’s hypothesis. Courts should not reject them out of hand, but should consider instead the information they provide about metabolic processes together with other proffered studies in assessing the admissibility of testimony based upon such studies.

d. *Structure-Activity Relationships*. Structure-activity relationship studies, which also rarely get admitted in court, explore the structure of the chemical at issue and how its structure is related to biological effects, by comparing it to other chemicals with similar

³²⁵ Cf. Beck et al., *supra* note 149, at 23 (explaining that in vitro studies are employed less frequently than in vivo animal studies). For a definition of in vivo and in vitro animal studies and an explanation of the difference between them, see, for example, *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1453 (D.V.I.) (“*In vivo* animal studies are those conducted in a living animal. *In vitro* animal studies are those employing animals [sic] cells in a controlled environment, such as a test tube or Petri dish.”), *aff’d*, 46 F.3d 1120 (3d Cir. 1994).

³²⁶ See Goldstein & Henifin, *supra* note 123, at 191 (discussing in vitro tests and criteria for their validity).

³²⁷ See Beck et al., *supra* note 149, at 23.

³²⁸ *Id.* at 25.

³²⁹ See, e.g., *Raynor v. Merrell Pharms., Inc.*, 104 F.3d 1371, 1374 (D.C. Cir. 1997) (affirming district court’s exclusion of plaintiffs’ experts who relied in part on in vitro studies); *Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 811, 813 (6th Cir. 1994) (affirming district court’s exclusion of expert testimony that did not establish connection between in vitro tests and plaintiffs’ illnesses). There is some reason for this reluctance. These tests rarely have been compared with outcomes in animal studies. See Goldstein & Henifin, *supra* note 123, at 191. Nonetheless, if an in vitro test consistently and accurately measures toxicity and predicts outcomes on the same cells or organs as in live studies, the results should not be ignored. See *id.* (listing criteria for reliability of in vitro tests).

structure and known toxic effects.³³⁰ Structure-activity relationships are a standard part of toxicology.³³¹ They too provide key insights into the ways chemicals affect the body.³³²

Both *in vitro* studies and structure-activity relationship studies illuminate the mechanisms of a chemical at the cellular or organ level.³³³ They are not inferior types of studies, as they present important information that needs to be accounted for in any valid hypothesis. Each of these kinds of studies is only a small piece of the puzzle; as many pieces as possible are needed to put the puzzle together. But any puzzle or theory is more or less incomplete, and the judge's job is to decide whether or not the pieces *could plausibly* fit together the way the expert says they do, or whether there are just too many pieces missing to tell.³³⁴ It is emphatically *not* the judge's job to decide whether the pieces *do* actually fit together the way the expert says they do. Rather, it is up to the judge to decide only whether the studies on which the expert relied—the pieces of the puzzle—provide plausible answers to questions of fact, not whether they provide the “correct” answers.³³⁵

³³⁰ See EPA Proposed Guidelines, *supra* note 10, at 17,977-78; Beck et al., *supra* note 149, at 25 (describing use of structure-activity relationship studies in toxicology); Richard Stone, *Zeroing In on Brain Toxins*, 255 *Science* 1063, 1063 (1992) (explaining that structure-activity relationship method involves comparing a chemical's structure to that of known toxin).

³³¹ See Goldstein & Henifin, *supra* note 123, at 216 (defining structure-activity relationships as “a method used by toxicologists”). The FDA, for example, requires studies of structural-activity relationships, among other studies, before a new drug will be approved. See 21 C.F.R. § 312.21(a)(2) (2000).

³³² For example, these relationships form a vital part of physiologically based pharmacokinetic modeling, the study of the “uptake and disposition of chemicals based on quantitative interrelationships among the critical biological determinants of these processes.” Krishnan & Anderson, *supra* note 311, at 149; see also *id.* at 158-61 (describing reactions of organic chemicals, nonvolatile organs, and inorganic chemicals).

³³³ See Beck et al., *supra* note 149, at 25 (“[S]tructure-activity relationships have been used to predict mutagenicity, lethality, and carcinogenicity. . . . [And *in vitro* tests may] support the classification of a chemical as a possible human carcinogen or the use of a linear dose-response model for carcinogenesis.”).

³³⁴ Thus, if the only information about toxicity were based on *in vitro* studies, it would be too difficult to see the whole picture, and the testimony should be inadmissible. But animal studies, *in vitro* studies, and structure-activity relationship studies together paint a sufficiently complete picture for the judge to assess validity. See Beck et al., *supra* note 149, at 25 (explaining that “there is no ‘best’ source of information”); Interdisciplinary Panel on Carcinogenicity, *supra* note 152, at 686 (noting that studies of various types must be examined together to provide reliable estimate of chemical's risk to humans).

³³⁵ See Conley & Peterson, *supra* note 9, at 1198 (arguing that while trial courts should determine whether “expert's methods yield answers to questions that are properly within the purview of the trier of fact,” whether they also should make “value judgment[s] about the soundness of those answers is a more controversial question”). Thus, the court in *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Or. 1996), was mistaken when it excluded expert testimony because the proffered studies failed to show that it was more

3. *Is Individualization a Valid Assumption?*

It should be clear at this point that living creatures share many attributes. What then of claims made by forensic scientists that they can achieve “absolute specificity and absolute identification” of individuals?³³⁶ Such a claim is based on an assumption that “nature never repeats.”³³⁷ Underlying the forensic sciences is the idea that people have unique fingerprints, handwriting, bitemarks, etc.³³⁸ Is the underlying assumption of these techniques—that people have unique characteristics—scientifically valid?

There is little empirical support for an assumption of uniqueness.³³⁹ Nor is there a testable theory that would explain the concept of individualization.³⁴⁰ Probability theory—on which such techniques purportedly are based—reflects the antithesis of absolute statements about uniqueness.³⁴¹ The unlikelihood of a coincidence of particular features must be based on the collection of data and the use of statisti-

likely than not that silicone caused the plaintiff's diseases. See *id.* at 1405. Deciding whether the expert's conclusion is relevant and admissible is not the same as deciding whether the conclusions are right. Assessing correctness is the province of the factfinder.

³³⁶ David A. Stoney, *What Made Us Ever Think We Could Individualize Using Statistics?*, 31 *J. Forensic Sci. Soc'y* 197, 197 (1991).

³³⁷ Michael J. Saks, Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science, 49 *Hastings L.J.* 1069, 1083 (1998) (quoting Harold Cummins & Charles Midlo, *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics* 150 (1943)).

³³⁸ See Beecher-Monas, *supra* note 3, at 57 (explaining that theory that these traits are uniquely personal has not undergone rigorous scientific testing necessary to meet standard of admissibility under *Daubert*).

³³⁹ See Saks, *supra* note 337, at 1083, 1085-86 (observing “almost complete lack of factual and statistical data pertaining to the problem of establishing identity” in forensic sciences and explaining that absence of frequent fingerprint matches does not establish their uniqueness any more than statement “all swans are white” establishes that all swans really are white just because no one has seen black one yet (internal quotation marks omitted)).

³⁴⁰ See *id.* at 1084 (“No articulated theory exists that explains why unique identifiability must be the order of the universe.”).

³⁴¹ See Brennan & Carter, *supra* note 84, at 46-47 (noting that “[p]robabilistic evidence deals openly with uncertainty” and rejecting notion that probabilistic evidence is insufficiently scientific); see also Michael J. Saks & Jonathan J. Koehler, *What DNA “Fingerprinting” Can Teach the Law About the Rest of Forensic Science*, 13 *Cardozo L. Rev.* 361, 361 (1991) (advocating probabilistic analysis of DNA typing as model for other forensic sciences). Saks explains that although the forensic scientists base their claims to individualization on “the multiplication rule of probability applied to populations,” Saks, *supra* note 337, at 1086, “probabilistic models cannot prove absolutes such as that no two are alike,” *id.* at 1087. Further, Saks notes that the assumption that the variables are independent may be questionable, and that the system of measurement and classification must itself be tested. See *id.* Moreover, he observes, fingerprint identification experts “use no probability models and have no probability data to use.” *Id.* at 1105-06.

cal analysis—requirements that have been startlingly absent from this branch of courtroom science.³⁴²

Even when empirical data exists, such data support only *qualified* possibilities of individualization. DNA testing explicitly relies on the presence of a database drawn from the general population for its ability to identify an individual.³⁴³ It also recognizes the implausibility of making an absolute identification statement.³⁴⁴

Of all the forensic identification techniques, DNA typing—originally developed not for the courtroom, but for research purposes—is the only technique that explicitly addresses and indeed depends on the probabilistic nature of the evidence.³⁴⁵ Thus, the assumption that “nature never repeats” is an assumption not grounded in valid science.³⁴⁶ Courts should not accept an assertion of individualization without empirical support.³⁴⁷ Rather, judges should adopt the alternative assumption, that individuals share many characteristics, as more scientifically sound in the absence of data to the contrary.

D. *Examine the Methodology*

Methodology refers to the way an experiment (or series of experiments) is conducted and encompasses the process of gathering, measuring, and reporting data.³⁴⁸ Once the court has filled gaps in knowledge with valid scientific default assumptions, the judge must examine the workings of the studies themselves to ensure that these assumptions may be employed fairly. The soundness of the methodology is what gives the data derived from that methodology its credibility.³⁴⁹ Evaluating study design, the data underlying each study, and

³⁴² See Beecher-Monas, *supra* note 3, at 84-85 (noting that although scientific evidence in criminal trials often involves probability statements, few lawyers or judges understand these statements’ statistical basis or underlying assumptions).

³⁴³ See, e.g., John Horgan, *High Profile: The Simpson Case Raises the Issue of DNA Reliability*, *Sci. Am.*, Oct. 1994, at 33, 36 (explaining frequency calculation based on frequency of genetic combination in general population, “and then multiply that frequency to obtain the odds of a random match”).

³⁴⁴ See *id.* (explaining that frequency calculation is merely estimate and that “[a] DNA profile is not as unique as, say, a fingerprint”).

³⁴⁵ See Saks, *supra* note 337, at 1083 (noting that DNA typing is only branch of forensic science that “takes the burdens of the probabilistic nature of forensic identification science seriously”).

³⁴⁶ Cf. *State v. Kunze*, 988 P.2d 977, 991-92 (Wash. Ct. App. 1999) (citing principle with general approval, but acknowledging that it might not be “susceptible of rigorous proof”).

³⁴⁷ As Saks explains, “the existence and nature of probability data are at the heart of the theory of forensic identification,” making the absence of both underlying data and statistical inferences drawn from them troubling. Saks, *supra* note 337, at 1124.

³⁴⁸ See Lu, *supra* note 304, at 84 (listing elements of “good laboratory practice” guidelines).

³⁴⁹ See John Ziman, *Reliable Knowledge: An Exploration of the Grounds of Belief in Science* 130-32 (1978) (discussing importance of methodology).

how that data was gathered, are important aspects of validity assessments.³⁵⁰

To test an articulated scientific hypothesis, the scientist must design an experiment: a series of observations calculated (provisionally) to corroborate (or falsify) the theory.³⁵¹ In designing an experiment, the researcher decides how large a sample must be used so that statistically significant results may be obtained.³⁵² In addition, the researcher attempts to control the experiment so that a minimum of variables affects the results or complicates the discovery of a proposed causal association.³⁵³ Whether or not a study validly shows a significant association between exposure to a given chemical and its effects in the body, for example, depends on the analysis of the data.³⁵⁴ Any useful description of the methods used in a study should include a detailed description of the data and the reasons for the selection of variables and control processes.³⁵⁵

³⁵⁰ See Thomas D. Cook & Donald T. Campbell, *Quasi-Experimentation: Design & Analysis Issues for Field Settings* 39 (1979) (stating that evaluation of any experiment requires critique of data and studies). In *Mancuso v. Consolidated Edison Co.*, 967 F. Supp. 1437 (S.D.N.Y. 1997), *aff'd in part and vacated in part on other grounds mem.*, 216 F.3d 1072 (2d Cir. 2000), for example, the proffered expert purported to rely on a series of studies but did not describe the studies in detail or attach abstracts of the studies to the expert's affidavit. The court rightly excluded the testimony. See *id.* at 1447. Thus, the importance of examining the details of experimental data in making admissibility determinations can hardly be overemphasized. For example, technician error, instrument error, laboratory procedural error, and errors in experimental technique all may invalidate a conclusion. See David L. Faigman et al., *Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence*, 15 *Cardozo L. Rev.* 1799, 1834 (1994) (reviewing data collection techniques, laboratory notebooks, and other evidence regarding "the specifics of the actual data produced for admission in court [that] must be examined").

³⁵¹ See Gad, *supra* note 81, at 21 (noting that experimentation has "a twofold purpose": "whether or not an agent results in an effect" and "how much of an effect is present"). However, as Popper pointed out, the scientist at most can demonstrate that the data provisionally corroborate the theory. See Popper, *supra* note 36, at 275.

³⁵² See Gad, *supra* note 81, at 24 (explaining that fundamental aspect of experimental design "is determining sufficient test and control group sizes to allow one to have an adequate level of confidence in the results of the study").

³⁵³ See *id.* at 21-22 (discussing importance of controls); Fienberg et al., *supra* note 82, at 17 (same); see also Knorr-Cetina, *supra* note 47, at 21 ("The major task of the laboratory is to rule out possibilities, manipulate the balance of choices so that one becomes more attractive than the others, and to up- or downgrade variables with respect to alternatives.").

³⁵⁴ See Gad, *supra* note 81, at 21-31 (discussing importance of experimental design and data analysis); Rothman, *supra* note 203, at 445 (explaining that because "[m]easurement is never exact[,] observations and the inferences drawn from them are subject to various errors," and therefore "[s]tatistical procedures have been developed in recognition of the need to evaluate measurement error"); see also EPA Proposed Guidelines, *supra* note 10, at 17,973-75 (discussing criteria for assessing adequacy of epidemiologic studies).

³⁵⁵ See EPA Proposed Guidelines, *supra* note 10, at 17,973 (listing "clear, well-documented, and appropriate methodology for data collection and analysis" and "complete and

In determining whether an experiment has yielded “reliable” data, a scientist looks at factors relating to the control exerted over the experiment.³⁵⁶ Every experiment should have “control” subjects to compare with the subjects of the experiment. One suspect practice that should set off warning bells for judges is the use of “historical controls,” control subjects based on previous experiments.³⁵⁷ Each time a study is performed, controls should be in place that account for extrinsic variables.³⁵⁸ Background “noise” that could affect the experiment must be eliminated as much as possible—the laboratory and its equipment must be clean, for example, to prevent contamination of the procedure—or taken into account, although inevitable human error means that no experiment ever can be performed perfectly.³⁵⁹ By performing a number of experiments and determining how far the re-

clear documentation of results” as two criteria for assessing adequacy of epidemiologic studies).

³⁵⁶ Reliability is the ability to obtain *consistent* results using a given experimental procedure. See Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, A Half-Century Later*, 80 Colum. L. Rev. 1197, 1201 n.20 (1980). Reliability should not be confused with validity. See *id.* (stating that validity, as opposed to reliability, entails test’s accuracy). A set of results under given conditions may be consistent—and thus reliable—but wrong (and therefore invalid). A good example of this phenomenon may be found in the hypothetical discussion of a new test for blood alcohol level that “grossly underestimates the amount of alcohol in one’s bloodstream, and yet . . . underestimates the blood alcohol level in one’s bloodstream *by the same amount every time.*” *Developments in the Law: Confronting the New Challenges of Scientific Evidence*, *supra* note 6, at 1534 (emphasis added). Such a test is reliable, but invalid. See Kaye & Freedman, *supra* note 83, at 341 (noting importance of reliability and distinguishing between reliability and accuracy).

³⁵⁷ The reason historical controls—controls from past experiments—are suspect is that they are not really controls. That is, the historically controlled subjects have not been subjected to exactly the same conditions as the test subjects. For a detailed explanation of why this is a problem, see Gad, *supra* note 81, at 22 (explaining that controls should be concurrent and “should come from the same source, lot, age, etc. as test group”); Interdisciplinary Panel on Carcinogenicity, *supra* note 152, at 683 (disfavoring historical controls). The court-appointed immunology expert in *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Or. 1996), rejected one of the studies relied on by the plaintiff’s expert for reasons including the study’s use of historical controls. See *id.* app. C at 1457-58. As one scientist has noted, “you can prove anything with selective controls,” so one must be wary of historical controls. Jon Cohen, *Cancer Vaccines Get a Shot in the Arm*, 262 *Science* 841, 843 (1993) (quoting scientist Charles Moertel).

³⁵⁸ “Generally speaking, the greater control a scientist has over the intervention being studied and the secondary variables, the greater the ability to make inferences that take account of the secondary variables without heroic assumptions. . . . But any method requires a model and, therefore, assumptions.” Fienberg et al., *supra* note 82, at 17.

³⁵⁹ Patrick Suppes calls the background noise problem “*ceteris paribus* conditions” and acknowledges that “[d]etailed information about the distribution of physical parameters characterizing the experimental environment is not a simple matter to incorporate in models of data and is usually not reported in the literature; roughly speaking, some general *ceteris paribus* conditions are assumed to hold.” Patrick Suppes, *Models of Data, in Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* 252, 258 (Ernest Nagel et al. eds., 1962).

sults deviate from each other and from those expected under the theory, the scientist determines the study's reliability.³⁶⁰

Principles of experimental design differ widely from field to field in science; some fields require more exacting protocols than others.³⁶¹ At a minimum, however, certain basic experimental design features must be either present or taken into account. The use of controls is one such basic design feature. For example, in toxicology studies, the experiment will be designed to hold all variables constant except for the exposure to the chemical in question.³⁶² The implicit assumption underlying the use of such controls is that any changes observed will be due to the chemical and not due to some other factor.³⁶³ Another basic design feature in toxicology experiments is the use of a range of doses. In a good study, a range of doses should be used in order to reveal the threshold level when a dose has no measurable effect.³⁶⁴

The design of an experiment, or a series of experiments, should incorporate standards of appropriate laboratory practice or provide a logical basis for disregarding them.³⁶⁵ In every field, the experiment

³⁶⁰ As noted above, reliability is an important concept in evaluating experiments. See *supra* note 356 and accompanying text. Essentially, this factor asks whether a scientific study can be or has been repeated. See Gad, *supra* note 81, at 21-22 (explaining importance of replication of results). Courts, presented with a limited number of studies upon which proffered experts rely, may have no way of knowing whether studies are repeatable. Courts, on the other hand, can determine whether a sufficient number of subjects was used to ensure accuracy in response measurement. See *id.* Generally, where reliability becomes a contested issue, courts have little trouble finding the study invalid. See, e.g., *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1462 (D.V.I.) (rejecting rabbit study as poorly done and unrepeatably), *aff'd*, 46 F.3d 1120 (3d Cir. 1994).

³⁶¹ Protocols offer a way of standardizing methodology and address such aspects of methodology as experimental conditions, reliability, and controls. Cf. Fienberg et al., *supra* note 82, at 17 (noting that "the greater control a scientist has over the intervention being studied and the secondary variables, the greater the ability to make inferences that take account of the secondary variables without heroic assumptions").

³⁶² See Lu, *supra* note 304, at 74-85 (discussing methodology for conventional toxicity studies).

³⁶³ See *id.*

³⁶⁴ See Rolf Hartung, *Dose-Response Relationships*, in *Toxic Substances and Human Risk*, *supra* note 255, at 29, 40 (explaining importance of experimental design).

³⁶⁵ For experiments in clinical medicine and drug testing, the FDA's Good Laboratory Practices and Good Clinical Practices provide the most-followed guidelines, and must be followed by studies supporting applications to the FDA. See *Good Laboratory Practice for Nonclinical Laboratory Studies*, 21 C.F.R. § 58 (2000); *Good Clinical Practice: Consolidated Guideline*, 62 Fed. Reg. 25,692 (1997); see also Foster & Huber, *supra* note 37, at 100-01 (noting that "the most widely used" guidelines to ensure reliable data "are Good Laboratory Practices (GLP) and Good Clinical Practices (GCP)"). GLP pertains to nonclinical laboratory studies conducted to support applications for research or marketing permits for products regulated by the FDA. GCP provides standards for clinical trials that involve human subjects. These formal guidelines are not applicable to basic scientific research, where the methods are new or have modified existing protocols; nonetheless, there are fundamental protocols in all fields, and departures from them should be explained. See *Good Clinical Practice: Consolidated Guideline*, 62 Fed. Reg. at 25,697. In the context of

must be designed so that the results honestly can be attributable to the data.³⁶⁶ Good methodology requires that the scientist develop and follow a protocol, use extensive quality control measures, and submit to outside audits of her studies.³⁶⁷

Courts in criminal cases have the most difficulty dealing with laboratory standards and protocols since criminal laboratories are largely unmonitored and unregulated.³⁶⁸ In civil toxic torts, on the other hand, courts frequently have gone to the opposite extreme, requiring strict adherence to conventional methodology.³⁶⁹ Although perfection is unachievable and straying from the standards may be necessary, the expert should be able to account for such deviation, giving adequate reasons for the variance.³⁷⁰ When the study relied upon employed sound statistical design, followed the protocols for such studies in the relevant field, observed proper control techniques, and demonstrated its reliability through proper techniques, the methodology is sound.

DNA testing, litigation and criminal prosecutions often have focused on laboratory controls. See, e.g., *State v. Bogan*, 905 P.2d 515, 521 (Ariz. Ct. App. 1995) (testing of seed pods found in defendant's truck and at crime scene made tests more reliable due to blind testing conditions); *People v. Marlow*, 41 Cal. Rptr. 2d 5, 17-18 (Ct. App.) (discussing, but rejecting, defense expert's complaints of lack of duplicate testing, lack of controls, failure to accurately measure results, and performance on proficiency tests), review granted, 899 P.2d 65 (Cal. 1995), review dismissed, 987 P.2d 695 (Cal. 1999); *People v. Simpson*, No. BA097211 (Cal. Super. Ct. Oct. 3, 1996).

³⁶⁶ See, e.g., Lu, *supra* note 304, at 57-67 (discussing factors that can modify toxic effects).

³⁶⁷ For examples of established methodological guidelines, see, e.g., Good Laboratory Practice for Nonclinical Laboratory Studies, 21 C.F.R. § 58 (2000); Good Clinical Practice Consolidated Guideline, 62 Fed. Reg. 25,692 (1997); AOAC Int'l, *Official Methods of Analysis of AOAC International* (Patricia Cunniff ed., 16th ed. 1995) (establishing protocols and methodologies for many different subjects of analysis).

³⁶⁸ See Beecher-Monas, *supra* note 3, at 67 (noting failure of forensic laboratories to follow appropriate scientific standards and practices).

³⁶⁹ For an example of a court's parsing of expert methodology, see *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441 (D.V.I.), *aff'd*, 46 F.3d 1120 (3d Cir. 1994), in which the court examined the inadequacies of a rabbit study relied on by plaintiff's expert, finding that, overall, too few rabbits were used to achieve statistical significance, each dosage group consisted of a single rabbit, there were virtually no controls, and the expert failed to include dead rabbits in the data. See *id.* at 1458-61. *Wade-Greaux* provides both a good example of the necessity of a civil court to examine thoroughly the details of a study's methodology, and an excellent analysis of the methodology, as the court revealed serious flaws in the proffered study. Most criminal courts do not examine methodology in any more than a cursory fashion. For a criminal case refusing to examine the methodology at all, see the analysis of *State v. Council*, 515 S.E.2d 508 (S.C.), *cert. denied*, 120 S. Ct. 588 (1999), in *infra* Part III.B.

³⁷⁰ See, e.g., *Lust v. Merrell Dow Pharms., Inc.*, 89 F.3d 594, 596-98 (9th Cir. 1996) (*affirming* district court's exclusion of expert's testimony because of failure to explain why key feature in teratology methodology, observation of distinctive pattern of birth defects, should not be required from studies upon which expert relied).

*E. Probabilistic Assessment of Expert Conclusions:
Putting It All Together*

The last step in the heuristic involves putting together answers to the four previous inquiries (into the expert's theory, empirical support, assumptions, and methodology) to resolve the admissibility question: Is the expert's conclusion based on sound science?³⁷¹ In essence, this is a two-part question: (1) Are the studies the expert relied upon relevant to the expert's conclusion (do they meet the requirement of "fit"), and (2) does the expert have corroborating empirical support for the proffered conclusion?³⁷² The first question of "fit" requires an inquiry into whether the proffered studies add empirical content to the hypothesis. Do they add a piece to the puzzle? Do they add anything to the expert's explanation of how things work? The second question of empirical support requires examining the variety and methodology of tests to which the hypothesis has been subjected, as well as those tests "that might have been performed but were not."³⁷³ The reason for any absence of testing also must be taken into account. For example, in toxic torts, chemical manufactur-

³⁷¹ The task for the judge in an admissibility determination is not to determine "whose evidence is most persuasive and exclude the evidence pointing the other way," but to decide whether the evidence reasonably supports the expert's hypothesis. Beecher-Monas, *supra* note 8, at 1073; see also Steve Gold, Note, Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence, 96 *Yale L.J.* 376, 385 (1986) (observing, pre-*Daubert*, that courts conflate burdens of proof and standards of persuasion); cf. Theresa M. Beiner, The Misuse of Summary Judgment in Hostile Environment Cases, 34 *Wake Forest L. Rev.* 71 (1999) (discussing misuse of summary judgment standard outside *Daubert* context). Conflating sufficiency with admissibility, although mistaken, still may continue in post-*Daubert* cases. See Thomas J. Mack, Scientific Testimony After *Daubert*: Some Early Returns from the Lower Courts, *Trial*, Aug. 1994, at 23, 30 ("[U]sing admissibility criteria in resolving sufficiency cases may not only continue but may do so with seeming validation from the *Daubert* opinion."). For cases conflating the standards, see *Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 814 (6th Cir. 1994) (assuming admissibility of plaintiffs' expert testimony but granting summary judgment to defendant since studies insufficient); *In re Hanford Nuclear Reservation Litig.*, No. CY-91-3015-AAM, 1998 WL 775340, at *139-*40 (E.D. Wash. Aug. 21, 1998); *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1398 (D. Or. 1996) (noting that plaintiffs must "prove not only the possibility of a causal connection between breast implants and the alleged systemic disease, but the medical probability of a causal connection" (emphasis added)); *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490, 1497 (E.D. Ark. 1996) (noting that issue of causation is "different" from issue of admissibility and can be reached only by jury and only after initial issue of admissibility is determined), *aff'd per curiam*, 133 F.3d 1132 (8th Cir. 1998).

³⁷² Only if the "scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true" may the judge grant summary judgment or direct a verdict. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 596 (1993).

³⁷³ Schum, *supra* note 146, at 243-44 (discussing Baconian probability method of Jonathan Cohen, and noting importance of both "the number of evidential tests a hypothesis survives and . . . the number of tests that might have been performed but were not").

ers have a strong disincentive to engage in safety research that may provide information that could later be used against them.³⁷⁴ Tests that would be helpful in establishing the causation hypothesis therefore may be absent, and this should not automatically weigh against admissibility of expert testimony, providing it has some empirical basis.³⁷⁵ In other contexts, however, absence of testing may indicate a flaw in the expert's reasoning. For example, in the criminal context, expert failure to test the theory of individualization severely undermines the strength of identification evidence.³⁷⁶

In assessing the expert's conclusion, it is important to understand the expert's language. Courts frequently misunderstand the language that scientists use to articulate their conclusions.³⁷⁷ This applies to the language of individual studies relied upon as well as to the expert's ultimate conclusion. For courts seeking certainty to justify their determinations, it is very frustrating to hear that a hypothesis only can be tested empirically, but that it never can be proven true.³⁷⁸ Unfortunately, however, there is no method of discovering the absolute truth in science.³⁷⁹ Indeed, there is no method of ascertaining whether a hypothesis is even probable.³⁸⁰ The most that a scientific expert can say is that a theory is consistent internally, consistent with other theories that have withstood repeated critique, and consistent with the data.³⁸¹ For a scientist to testify to more is intellectually dishonest. What science offers is not the truth, not certainty, but explanation.³⁸²

In addition to the inability of an intellectually honest scientist to say whether a proffered hypothesis is "probable," the language of scientific conclusions often troubles judges. Because scientists seek to

³⁷⁴ Cf. *supra* notes 322-24 and accompanying text (explaining why burden should be placed on manufacturers to prove when there is no equivalence in exposure routes).

³⁷⁵ See Beecher-Monas, *supra* note 8, at 1090-91 (describing how manufacturers resist performing research that would yield evidence that could be used against them).

³⁷⁶ See *supra* Part II.C.3 (discussing problems with theory of individualization).

³⁷⁷ See, e.g., *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1320 (11th Cir. 1999) (requiring expert to testify in terms of probability of causation).

³⁷⁸ See *supra* notes 74-76 and accompanying text (discussing Popperian notion that scientists cannot establish truth, but only probabilities).

³⁷⁹ See *supra* note 36 (discussing Popper's view of science as uncertain).

³⁸⁰ See Popper, *supra* note 36, at 317 (observing that science is "a system of guesses or anticipations . . . with which we work as long as they stand up to tests, and of which we are never justified in saying that we know that they are 'true' or 'more or less certain' or even probable").

³⁸¹ See *id.* at 32-33. Science constantly must question and criticize its theories, even those that turn out to be true. See Beecher-Monas & Garcia-Rill, *supra* note 5, at 254-55 (discussing Popper's philosophy of science).

³⁸² See Heidi Li Feldman, *Science and Uncertainty in Mass Exposure Litigation*, 74 *Tex. L. Rev.* 1, 17 (1995) (proposing that "science does not produce fixed, unassailable conclusions" but rather that "uncertainty among scientists is a natural state of affairs"); *supra* note 36 (discussing Popper's view of science as uncertain).

exclude false positives—to avoid claiming there is an effect when there is none—scientific conclusions are couched in tentative phrases.³⁸³ “Association” is preferred to “causation,” for example.³⁸⁴ Thus, failing to understand that causation, like other hypotheses, can never be proven true, courts even may reject as unreliable evidence that easily meets scientific criteria for validity.

It is equally nonsensical for judges to discard a study just because it reaches a different conclusion than does the testifying expert. Experts can draw diametrically opposed conclusions about a perfectly valid study.³⁸⁵ That does not mean that one of the experts has drawn an invalid conclusion or that any conclusion would be speculative. It is more likely that the experts have used different underlying assumptions to fill the information gaps.³⁸⁶ This is in no way illegitimate, or even unusual. As long as the study’s data support the expert’s hypothesis, the expert justifiably can rely on the study for support. Thus, the expert’s *reasoning*, rather than the conclusions of the various studies relied on, should be the focus of the inquiry. Unfortunately, because they fail to understand the language of scientific conclusion drawing, courts frequently exclude expert testimony on this basis.³⁸⁷

³⁸³ See, e.g., Letter from Merwyn R. Greenlick to U.S. District Judge Robert Jones and Oregon Court of Appeals Judge Nely Johnson, reprinted in *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 app. B at 1447 (D. Or. 1996) (reporting that court-appointed expert explained that “in epidemiological science the odds are stacked against incorrectly rejecting the ‘null hypothesis’ . . . that there is no relationship between the suspected causal factor and the disease”).

³⁸⁴ See *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 727 (Tex. 1997) (noting that “the discipline of epidemiology studies associations, not ‘causation’ per se”); cf. *Bailey et al.*, supra note 39, at 157 (observing that “[m]ost researchers are conservative when it comes to assessing causal relationships, often calling for stronger evidence and more research before a conclusion of causation is drawn”).

³⁸⁵ See supra notes 164-66 and accompanying text.

³⁸⁶ See, e.g., 1 Schum, supra note 16, at 17 (explaining that “[t]wo astute, coherent, and knowledgeable analysts, having access to the same evidence[] [may] reach entirely different conclusions . . . [because] their initial premises, hypotheses, or expectations may differ”).

³⁸⁷ For example, in *General Electric Co. v. Joiner*, 522 U.S. 136 (1997), the plaintiff proffered a study of workers at an Italian capacitor plant who had been exposed to PCBs to support the plaintiff’s theory that exposure to PCBs caused cancer. See *id.* at 145. The study’s authors “noted that lung cancer deaths among ex-employees at the plant were higher than might have been expected” but ultimately concluded that “there were apparently no grounds for associating lung cancer deaths . . . and exposure in the plant.” *Id.* The Court rejected this study since its authors reached a different conclusion than did the testifying expert who relied on the study’s data rather than its conclusions. See *id.* (“Given that [the study’s authors] were unwilling to say that PCB exposure had caused cancer among the workers they examined, their study did not support the experts’ conclusion that Joiner’s exposure to PCB’s caused his cancer.”); see also, e.g., *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490, 1517 (E.D. Ark. 1996) (excluding expert testimony based on studies that stated opinions that did not “go to the same extent as the opinions [expert] would put before the jury”), *aff’d per curiam*, 133 F.3d 1132 (8th Cir. 1998);

Another problem posed by the language of scientific conclusion drawing is that many—if not most—published studies conclude that more research needs to be done.³⁸⁸ Judges often reject experts relying on such studies, finding the studies irrelevant to the expert's testimony, and thus failing to meet the "fit" requirement, because they are too speculative.³⁸⁹ This judicial reaction reflects ignorance of the research enterprise. It is not surprising that scientists publishing the results of their studies in scientific journals conclude that more research needs to be done.³⁹⁰ Of course it does.³⁹¹ There are always gaps in our knowledge, and scientists are anxious to pursue the quest of filling them. How else to continue adding to the communal puzzle? How else to convince funding authorities to continue to support the effort? Good research tends to generate more research, as long as the funding can be found. It is thus ludicrous for a judge to exclude testimony merely because a study relied upon concludes that more research needs to be done. What is important is not the conclusion but the data.

So what makes an explanation scientifically sound? High empirical content is important.³⁹² But given the unavoidable gaps in our knowledge, how much empirical content should a judge demand for admissibility? There are a number of theories about evaluating evidence under uncertainty, each contributing important insights.³⁹³

The following inquiries may prove helpful to judges in deciding whether there is sufficient empirical support to admit an expert's theory into evidence. First, for each proffered study, how valuable is each study on its own? What piece of the hypothetical puzzle does

Rutigliano v. Valley Bus. Forms, 929 F. Supp. 779, 785 (D.N.J. 1996) ("The case law warns against use of medical literature to draw conclusions not drawn in the literature itself."), *aff'd mem.*, 118 F.3d 1577 (3d Cir. 1997).

³⁸⁸ See, e.g., Kelley v. American Heyer-Schulte Corp., 957 F. Supp. 873, 877 (W.D. Tex. 1997) (excluding report suggesting that "further research was necessary before firm conclusions could be drawn").

³⁸⁹ See *Joiner*, 522 U.S. at 144-45.

³⁹⁰ See, e.g., Sherine E. Gabriel et al., Risk of Connective Tissue Diseases and Other Disorders After Breast Implantation, 330 N. Eng. J. Med. 1697, 1701 (1994) (finding no statistically significant correlation between incidence of disease and breast implants, but noting that there are other "ongoing" studies whose "results, together with ours, will help to resolve this controversy").

³⁹¹ See A.F. Chalmers, *What Is This Thing Called Science?* 125-26 (2d ed. 1982) (observing that "objective opportunities" for further exploration are created by work of other scientists).

³⁹² See Popper, *supra* note 36, at 120-21 (arguing for importance of empirical corroboration).

³⁹³ See Peter Tillers, Mapping Inferential Domains, 66 B.U. L. Rev. 883, 887-88 (1986) (arguing that Bayesianism, Baconian rationalism, fuzzy set theory, and "scenario" theory are complementary systems and that "diverse models of inference are both permissible and necessary").

each study provide? How sound is the methodology underlying each study? What are the areas of uncertainty that the study fails to address? Given the knowledge of the first study (or set of studies), how much more support does the second study (or set of studies) add to the expert's hypothesis? What information gaps does the second study (or set of studies) fill? How much does the second study (or set of studies) explain in the absence of the first study (or set of studies)? How rare or unexpected are the results of each of the studies?³⁹⁴ The judge must consider the extent and strength of empirical support; the level of consistency within the theory, with other current theories, and with the data; the acceptability of underlying assumptions; whether each of the proffered studies is methodologically sound; and whether each contributes toward a biologically plausible theory.³⁹⁵

Because of the uncertainties inherent in scientific studies, studies that conclude that more research must be done or that reach different conclusions from that of the expert should not be considered illegitimate.³⁹⁶ Rather, such uncertainties compel the scientist (and the judge) to examine "how well all of the available evidence fits together with the underlying theory."³⁹⁷ In sum, in order to assess the scientific basis of the expert's conclusion, it is necessary to examine theory, data, assumptions, and methodology. The judge must ask: How well do they fit together? How good is the explanation they provide?³⁹⁸ No algorithm for inference can substitute for thinking carefully about the problem.³⁹⁹ A hefty dose of critical judgment is needed for this

³⁹⁴ See 2 Schum, *supra* note 16, at 192-93 (suggesting questions may be helpful in evaluating inferences drawn from evidence). While the unexpectedness of results may lead one to question methodology, if the methodology and assumptions are sound, unexpectedness in fact may signal a robust theory. See Popper, *supra* note 96, at 58 ("[E]very interesting and powerful statement must have a low probability . . .").

³⁹⁵ For more suggestions of questions to ask in assessing scientific evidence, see 1 Schum, *supra* note 16, at 124-33. Schum asserts that proper conclusion-drawing consists of answering six questions: "What Did You expect?," "From The Evidence, What Do You Know?," "Believable Evidence: What Does It Mean?," "How Consistent Is The Evidence?," "Can The Value Of Any Evidence Be Enhanced or Suppressed?," "What's Missing?" *Id.* at 124-27.

³⁹⁶ See Brennan & Carter, *supra* note 84, at 47 (stating that uncertainty "makes the theories which guide the research explicit, and inform[s] the raw conclusions gathered by scientists. . . . [I]t is a mistake—based on a mistaken view of the scientific enterprise—to repudiate such [probabilistic] evidence . . .").

³⁹⁷ Beecher-Monas, *supra* note 8, at 1094.

³⁹⁸ See Popper, *supra* note 36, at 32-33 (discussing how theories are tested by comparing conclusions drawn from them with empirical observation).

³⁹⁹ See Rothman, *supra* note 203, at 446-47 (urging replacement of significance tests with confidence intervals, but cautioning that even confidence intervals cannot be used as surrogate test "to determine whether the null hypothesis value falls inside or outside the confidence interval").

determination, but it should be informed and guided by an understanding of the scientific process and scientific language.

III

APPLYING THE HEURISTIC

Armed with the proposed heuristic, what would judges do differently? The following section examines the expert evidence in two cases in which the heuristic would have made a dramatic difference both in rationale and in result: one civil toxic tort case and one criminal identification case. In the toxic tort case, *Wright v. Willamette Industries, Inc.*,⁴⁰⁰ the Eighth Circuit overturned a jury verdict for plaintiffs who claimed their injuries were caused by the formaldehyde-laced emissions of a fiberboard factory because the court decided that plaintiffs' expert's testimony was not based on scientific knowledge.⁴⁰¹ *Wright* has been cited widely by other judges as a rationale for excluding expert testimony, both in the Eighth Circuit and elsewhere.⁴⁰²

The criminal case *State v. Council*⁴⁰³ is a potentially influential state case employing a *Daubert*-like standard to determine admissibility of mitochondrial deoxyribonucleic acid (mtDNA) testimony. *Council* was only the second case in which an expert proffered mtDNA testimony.⁴⁰⁴ Despite the novelty of such testimony, the court found no defense expert to be necessary, and without examining the prosecution expert's methodology or the basis for the expert's assertion of reliability, the court admitted the testimony in its entirety.

⁴⁰⁰ 91 F.3d 1105 (8th Cir. 1996).

⁴⁰¹ See id. at 1108.

⁴⁰² See, e.g., *Mitchell v. Gencorp Inc.*, 165 F.3d 778, 781 (10th Cir. 1999) (citing *Wright* for requiring proof of level of exposure and stating that plaintiffs fell short in their attempt to demonstrate actual level of exposure); *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 198 (5th Cir. 1996) (citing *Wright* in rejecting "weight of the evidence" methodology adopted by plaintiffs' experts); *National Bank of Commerce v. Associated Milk Producers, Inc.*, 22 F. Supp. 2d 942, 961-62 (E.D. Ark. 1998) (citing *Wright* and excluding causation testimony for lack of exposure level data), aff'd, 191 F.3d 858 (8th Cir. 1999); *Mancuso v. Consolidated Edison Co.*, 967 F. Supp. 1437, 1445 (S.D.N.Y. 1997) (citing *Wright* in noting widespread requirement that expert establish "general causation" through exposure level data in scientific literature), aff'd in part and vacated in part on other grounds, 216 F.3d 1072 (2d Cir. 2000).

⁴⁰³ 515 S.E.2d 508 (S.C.), cert. denied, 120 S. Ct. 588 (1999).

⁴⁰⁴ See *infra* note 480. The first time an expert proffered mtDNA testimony was in a Tennessee state court in August, 1996, shortly after this method of DNA testing was first implemented in the FBI Laboratory in June of 1996. See *State v. Ware*, No. 03C01-9705CR00164, 1999 WL 233592, at *13 (Tenn. Crim. App. Apr. 20, 1999). The defendant was convicted of murder and rape based partially on hair matching evidence using mtDNA testing. See id. On appeal, the defendant argued that the evidence was not yet scientifically reliable and should not have been admitted, but the appellate court held the admission of the evidence not to be an abuse of discretion, and that even if it was, the error was harmless. See id. at *14, *17.

The heuristic proposed in this Article would have altered the admissibility determination significantly in both of these cases.

A. *Wright v. Willamette Industries, Inc.*

In this toxic tort case, the plaintiffs were a family who lived a short distance from a fiberboard manufacturing plant, and suffered from headaches, sore throats, respiratory ailments, and dizziness, which they claimed were caused by emissions from the plant.⁴⁰⁵ It was undisputed that the plant emitted particles laced with formaldehyde,⁴⁰⁶ that the polluting emissions exceeded state maximum levels,⁴⁰⁷ and that the plaintiffs were exposed to these particles.⁴⁰⁸

After a jury trial at which the plaintiffs prevailed, they were awarded \$226,250 in compensatory damages.⁴⁰⁹ On appeal, the Eighth Circuit overturned the jury verdict, ruling that the plaintiffs were unable to meet their burden of proving proximate cause because their expert testimony should have been excluded as "speculation."⁴¹⁰ As far as the Eighth Circuit was concerned, although the plaintiffs demonstrated exposure to fiberboard particles produced by the defendant manufacturer, and that these particles were found "in their house, their sputum, and their urine, they failed to produce evidence that they were exposed to a hazardous level of formaldehyde from the fibers emanating from Willamette's plant."⁴¹¹ As support, the court cited the Federal Judicial Center's Reference Manual on "fit,"⁴¹² apparently for the notion that plaintiffs need to "prove adequate exposure to a toxic substance that was somehow connected to the defendant."⁴¹³ The Eighth Circuit's ruling stemmed from a fundamental misunderstanding of basic scientific theory, namely elementary principles of metabolism and threshold response.⁴¹⁴

1. *What's the Theory?*

In *Wright*, the plaintiffs' causation hypothesis was based on the uncontroversial theory that formaldehyde causes respiratory disease

⁴⁰⁵ See *Wright*, 91 F.3d at 1106.

⁴⁰⁶ See *id.* at 1106.

⁴⁰⁷ See *id.* at 1108 (Heaney, J., dissenting).

⁴⁰⁸ See *id.* at 1107.

⁴⁰⁹ See *id.* at 1106.

⁴¹⁰ *Id.* at 1108. For a discussion of the *Wright* court's remarkably undeferential standard of review, see *infra* note 468.

⁴¹¹ *Wright*, 91 F.3d at 1107.

⁴¹² *Id.* at 1108 (citing Federal Judicial Ctr. Manual, *supra* note 9, at 47-48).

⁴¹³ Federal Judicial Ctr. Manual, *supra* note 9, at 48.

⁴¹⁴ See *supra* Part II.C.2.b (discussing metabolism and threshold responses).

in humans.⁴¹⁵ More specifically, their experts' hypothesis proposed that when plaintiffs inhaled the formaldehyde contained on the fiberboard particles invading plaintiffs' home, the exposure led to the plaintiffs' particular array of symptoms. One expert testified that "the Wrights' complaints were more probably than not related to formaldehyde."⁴¹⁶ This hypothesis depended on testimony advanced by two other plaintiff experts that formaldehyde attached to fiberboard particles has similar respiratory effects to *gaseous* formaldehyde.⁴¹⁷ But this leap was too great for the Eighth Circuit. Thus, although the court could accept the theory that "gaseous formaldehyde" could cause human respiratory ailments, it found causation from formaldehyde attached to fiberboard particles to be unsubstantiated.⁴¹⁸

2. *Examine All the Available Evidence*

Although formaldehyde is ubiquitous in the civilized world,⁴¹⁹ the defendant plant's emissions far exceeded statutorily permitted levels.⁴²⁰ The manufacturer had failed to install available equipment to control the emissions.⁴²¹ Emissions from the plant "fell like 'snow'" on plaintiffs' property.⁴²² The manufacturer was the only source of formaldehyde-coated wood particles in the vicinity of the plaintiffs' home.⁴²³ Significant levels of the plant's toxic emissions were found in the plaintiffs' bodily fluids.⁴²⁴ In fact, the *Wright* plaintiffs were exposed to enough formaldehyde to cause metabolic by-

⁴¹⁵ See Robert E. Gosselin et al., *Clinical Toxicology of Commercial Products: Acute Poisoning* 166-67 (4th ed. 1976) (noting common symptoms from formaldehyde exposure including dizziness, nausea, skin and eye irritation, nose and throat irritation, and difficulty breathing).

⁴¹⁶ *Wright*, 91 F.3d at 1108 (assessing testimony of Dr. Frank Peretti); see also Appellees' Brief at 27, *Wright* (No. 95-4227).

⁴¹⁷ See *Wright*, 91 F.3d at 1107 (assessing testimony of Drs. Fred Fowler and Jimmie Valentine).

⁴¹⁸ See *id.*

⁴¹⁹ See Marian C. Marbury & Robert A. Krieger, *Formaldehyde*, in *Indoor Air Pollution: A Health Perspective* 223, 223-25 (Jonathan M. Samet & John D. Spengler eds., 1991) (discussing scientific and regulatory controversies arising from numerous reports published in late 1970s on "ubiquity of formaldehyde in consumer products and residences").

⁴²⁰ See Appellees' Brief at 24, *Wright* (No. 95-4227).

⁴²¹ See *Wright*, 91 F.3d at 1108 (Heaney, J., dissenting).

⁴²² *Id.* at 1109 (Heaney, J., dissenting).

⁴²³ See *Wright v. Willamette Indus., Inc.*, No. 94-6081 at 5 (W.D. Ark. Dec. 13, 1995) (unpublished order on file with the *New York University Law Review*).

⁴²⁴ See *Wright*, 91 F.3d at 1109 (Heaney, J., dissenting); Appellees' Brief at 17, *Wright* (No. 95-4227).

products to appear in their body fluids.⁴²⁵ None of this evidence was disputed.⁴²⁶

The plaintiffs' testifying experts included a forensic pathologist, a toxicologist, and a specialist in medical particulate research.⁴²⁷ The toxicologist and forensic pathologist testified to general causation.⁴²⁸ The particle science expert "demonstrated the mechanism by which the chemically coated particles entered the respiratory tracts of the [Wrights]."⁴²⁹ The forensic pathologist testified that this mechanism was similar to particles in the gaseous state.⁴³⁰ The plaintiffs' treating physician made a differential diagnosis that the plaintiffs' symptoms were "more probably than not related to their exposure to the plant emissions."⁴³¹ Each of the experts relied on government studies showing formaldehyde to be a carcinogen and respiratory irritant and to cause symptoms similar to those the plaintiffs suffered.⁴³²

There is ample evidence upon which the experts also presumably relied, although the court does not discuss it, that "[f]ormaldehyde can produce health effects ranging from acute nausea, eye irritation, and respiratory impairment to longer term effects, like cancer."⁴³³ Formaldehyde is a well-known allergen, as well as a mucous membrane irritant,⁴³⁴ causing "irritation to the eyes, nose, and throat, respiratory

⁴²⁵ See Appellees' Brief at 17, *Wright* (No. 95-4227).

⁴²⁶ See *Wright*, 91 F.3d at 1108-09 (Heaney, J., dissenting).

⁴²⁷ See Appellees' Brief at 25-26, *Wright* (No. 95-4227) (stating testimony of Drs. Jimmie Valentine, Malay Mazumder, and Frank Peretti).

⁴²⁸ See *Wright*, 91 F.3d at 1107; Appellees' Brief at 26, *Wright* (No. 95-4227).

⁴²⁹ Appellees' Brief at 25, *Wright* (No. 95-4227).

⁴³⁰ See *Wright*, 91 F.3d at 1108; Appellees' Brief at 26-28, *Wright* (No. 95-4227) (stating testimony of Dr. Frank Peretti).

⁴³¹ *Wright*, 91 F.3d at 1109 (Heaney, J., dissenting).

⁴³² See *id.*

⁴³³ 3 Frank P. Grad, *Treatise on Environmental Law* § 4A.01[4], at 4A-17 (2000). In 1979, the Chemical Industry Institute of Toxicology issued a report concluding that inhalation of formaldehyde causes nasal cancer in rodents. See William D. Kerns et al., *The Chronic Effects of Formaldehyde Inhalation in Rats and Mice: A Preliminary Report, in Formaldehyde Toxicity* 111-31 (James E. Gibson ed., 1983).

⁴³⁴ See Edward P. Horvath, Jr. et al., *Effects of Formaldehyde on the Mucous Membranes and Lungs: A Study of an Industrial Population*, 259 *JAMA* 701, 701 (1988) ("Airborne formaldehyde is a known irritant to the eyes, upper airways of the nose and throat, and lower airways of the lungs."); Thomas J. Kulle et al., *Formaldehyde Dose-Response in Healthy Nonsmokers*, 37 *Int'l J. Air Pollution Control & Hazardous Waste Mgmt.* 919, 919 (1987) ("Irritation of the eyes and upper respiratory tract is the most frequent finding associated with [formaldehyde] exposures."); Larry R. Sauder et al., *Acute Pulmonary Response to Formaldehyde Exposure in Healthy Nonsmokers*, 28 *J. Occupational Med.* 420, 420 (1986) ("Formaldehyde (HCHO) has been implicated as a cause of asthma-like symptoms in some individuals . . ."); E.N. Schachter et al., *A Study of Respiratory Effects from Exposure to 2 ppm Formaldehyde in Healthy Subjects*, 41 *Archives Envtl. Health* 229 (1986) ("Descriptions of the respiratory effects of formaldehyde have been primarily associated with irritative effects of this gas on the mucous membranes of the upper airways. . . .

illnesses, skin irritations, dizziness, headaches, and nausea."⁴³⁵ Exposure levels of 0.1 parts per million (ppm) may cause difficulty in breathing and asthma-like respiratory illnesses.⁴³⁶ The Occupational Safety and Health Administration's formaldehyde exposure standard is 0.75 ppm for an eight-hour period, with a maximum of 2 ppm for a fifteen-minute short-term exposure limit.⁴³⁷ Just having large amounts of particle board itself in the home is a risk factor for asthma, headache, and throat irritation.⁴³⁸ Moreover, the absorption rate of inhaled or ingested formaldehyde is very high.⁴³⁹ The EPA found that ninety-five percent of people respond to concentrations of 0.1 to 3 ppm.⁴⁴⁰ The plaintiffs' toxicologist expert drew on these findings in testifying that formaldehyde metabolites such as those found in the plaintiffs' urine "are known to be toxic in animals and to cause cell membrane disruption in humans."⁴⁴¹

3. *Use Valid Assumptions to Fill the Gaps*

The court found that the plaintiffs' experts' inability to demonstrate plaintiffs' exposure levels was dispositive and necessitated exclusion of the expert testimony and granting of judgment as a matter of law to the defendants.⁴⁴² There is no question that the plaintiffs

[A] number of reports suggest that asthmatic symptoms may follow formaldehyde exposure.").

⁴³⁵ Andrew N. Davis & Paul E. Schaffman, *The Home Environmental Sourcebook* 89 (1st ed. 1996).

⁴³⁶ See *id.* Whether the experts relied on these known levels is another question. Presumably, they did. If not, such ignorance constitutes an expert failure, and ultimately a lawyering failure, which is beyond the scope of this Article. This kind of failure can be remedied by the judge asking for supplemental briefing on the underlying studies on which the expert relied. Here, there is ample evidence available (and easily accessible) about formaldehyde and its effects. Indeed, it has been so well litigated, and the subject of numerous reports and publicized proceedings, that the court could have taken judicial notice of the harmful exposure levels. Moreover, the Wrights' expert toxicologist proffered testimony about the levels of exposure that would cause symptoms like the plaintiffs'. See *Wright*, 91 F.3d at 1107.

⁴³⁷ See *Occupational Exposure to Formaldehyde*, 57 Fed. Reg. 22,290 (1992) (codified at 29 C.F.R. § 1910.1048(c) (1999)).

⁴³⁸ See Peder Daugbjerg, *Is Particleboard in the Home Detrimental to Health?*, 48 *Envtl. Res.* 154 (1989).

⁴³⁹ See, e.g., John L. Egle, Jr., *Retention of Inhaled Formaldehyde, Propionaldehyde, and Acrolein in the Dog*, 25 *Archives Env'tl. Health* 119, 121 (1972) (describing experiment exposing mongrel dogs to various concentrations of aldehyde compounds and noting that "[u]ptake of formaldehyde remained near 100% regardless of the concentration").

⁴⁴⁰ See Office of Pesticides & Toxic Substances, EPA, *Assessment of Health Risks to Garment Workers and Certain Home Residents from Exposure to Formaldehyde* (1987).

⁴⁴¹ Appellees' Brief at 26, *Wright v. Willamette Indus., Inc.*, 91 F.3d 1105 (10th Cir. 1996) (No. 95-4227) (testimony of Dr. Valentine).

⁴⁴² The court stated that

[Dr. Peretti's] opinion was not based on any knowledge about what amounts of wood fibers impregnated with formaldehyde involve an appreciable risk of

were exposed to formaldehyde and that the formaldehyde to which they were exposed came from the defendant's lumber mill. The data showed a strong association between exposure to formaldehyde at low levels and symptoms such as the plaintiffs suffered. But the court concluded that because the plaintiffs presented no exposure level data, the rest of the testimony was irrelevant.⁴⁴³ In addition, the court was unconvinced that the absorption of gaseous formaldehyde leads to effects similar to exposure to formaldehyde bonded to particles of fiberboard. Thus, the court was unable to fill the two information gaps it faced: gaps as to exposure level and absorption mechanisms. Had it used the scientifically acceptable assumptions proposed in the heuristic, the court would not have had such difficulty.⁴⁴⁴

a. Exposure Level. The *Wright* court insisted that "a plaintiff in a toxic tort case must prove the levels of exposure that are hazardous to human beings generally as well as the plaintiff's actual level of exposure to the defendant's toxic substance" in order to recover.⁴⁴⁵ There is ample data to show the level at which formaldehyde affects people.⁴⁴⁶ The problem for the plaintiffs was that formaldehyde is

harm to human beings who breathe them. The trial court should therefore have excluded Dr. Peretti's testimony

. . . Without proving hazardous levels of exposure to Willamette's formaldehyde, the Wrights failed to carry their burden of proof at trial on the issue of causation

Wright, 91 F.3d at 1108.

⁴⁴³ See *id.*

⁴⁴⁴ *National Bank of Commerce v. Associated Milk Producers, Inc.*, 22 F. Supp. 2d 942 (E.D. Ark. 1998), *aff'd*, 191 F.3d 858 (8th Cir. 1999), is instructive, since the court correctly excluded causation testimony, but for the wrong reasons. The plaintiff in *Associated Milk Producers* claimed his laryngeal cancer had been caused by exposure to aflatoxins in contaminated milk at his workplace. The court excluded the testimony because the plaintiff was unable to show level of exposure. See *id.* at 981-82. Had the court used the default assumptions proposed in this Article, however, it would have recognized that a linear-dose response default makes knowledge of exposure levels unnecessary. A linear default is strongly defensible in cancer cases such as *Associated Milk Producers*. See *supra* notes 271-94 and accompanying text. The district court was unconvinced, citing Paracelsus's maxim that "the dose makes the poison" and disregarding the expert's testimony regarding linearity. *Id.* at 958. Using a linear dose-response assumption, the amount of contaminated milk to which the plaintiff was exposed would be irrelevant because exposure to any amount would pose a small, but not zero, risk of genetic damage ultimately leading to cancer. However, aflatoxin, even in its most virulent form—commonly found in corn and peanut butter—causes *liver* cancer rather than *laryngeal* cancer. See *id.* at 962; Bruce N. Ames et al., *Ranking Possible Carcinogenic Hazards*, 236 *Science* 271, 272-73 (1987). The default criteria for target organs suggested earlier would preclude using evidence of liver cancer causation to show causation for laryngeal cancer without a showing that aflatoxin affects a range of target organs. See *supra* notes 295-301 and accompanying text. Thus, the judge was correct in excluding the testimony but based his decision on the wrong reason.

⁴⁴⁵ *Wright*, 91 F.3d at 1106.

⁴⁴⁶ See *supra* note 433 and accompanying text.

ubiquitous and that they could not show exposure levels at a greater level than background.⁴⁴⁷ This does not mean, however, that they were not exposed to levels greater than background. A sample of emissions from the plant showed 18,000 to 20,000 ppm formaldehyde, quantities far in excess of state maximum levels.⁴⁴⁸ Although the plaintiffs' home was located at some distance from the lumber mill, it was close enough to receive particulate matter falling "like 'snow'" on the premises.⁴⁴⁹ Their bodily fluids had significant levels of formaldehyde and its metabolites. This is important because when the metabolite level is known, the dose can be calculated mathematically.⁴⁵⁰

Because formaldehyde is known to have a nonlinear response (that is, an exposure threshold),⁴⁵¹ the defendant lumber mill argued that the plaintiffs had to show exposure above this threshold level.⁴⁵² This argument, accepted by the Eighth Circuit, reflects a misunderstanding of the probabilistic nature of the threshold response. The Wrights were already showing symptoms consistent with exposure. They were already showing that formaldehyde had been absorbed and distributed throughout their bodies. In other words, whether or not most people would show symptoms at that level, *their* particular, individual threshold had been reached.

Threshold is a probabilistic concept stating that most people will not be affected until exposure exceeds a certain amount. The precise amount of chemical necessary to achieve a response may vary among individuals.⁴⁵³ Scientists understand that some individuals will be more sensitive to certain agents, and for them the threshold exposure level will be lower.⁴⁵⁴ Once the threshold has been reached, the response can be assumed to be linear.⁴⁵⁵ These plaintiffs demonstrated

⁴⁴⁷ See Appellant's Opening Brief at 18, *Wright* (No. 95-4227); Appellant's Reply Brief at 5, *Wright* (No. 95-4227). See supra note 191 for a definition of background levels of exposure. Everyone in industrialized countries is exposed to formaldehyde; this common exposure is the general background level. The Wrights, however, were exposed to more than the background level but could not prove precisely how much.

⁴⁴⁸ See *Wright*, 91 F.3d at 1108 (Heaney, J., dissenting) (noting that "the levels of formaldehyde emitted from the plant exceeded levels permitted by industry and state standards"); Appellee's Brief at 26, *Wright* (No. 95-4227) (stating testimony of Dr. Fred Fowler regarding level of emissions from Willamette plant).

⁴⁴⁹ *Wright*, 91 F.3d at 1108-09 (Heaney, J., dissenting).

⁴⁵⁰ See supra notes 302-07 and accompanying text.

⁴⁵¹ An exposure threshold is the level at which effects become observable. For a discussion of linear versus threshold responses, see supra notes 279-94 and accompanying text.

⁴⁵² See *Wright*, 91 F.3d at 1106.

⁴⁵³ See supra Part II.C.2.b.ii.

⁴⁵⁴ See Rees & Hattis, supra note 279, at 277.

⁴⁵⁵ Cf. Gad, supra note 81, at 235 (summarizing argument for existence of threshold that states that "[m]ost (if not all) carcinogens and mutagens exhibit a dose response relationship, resulting in an apparent or effective threshold for at least some agents").

exposure and symptoms consistent with exposure.⁴⁵⁶ There was testimony that “distinguished the effects of dust and other nontoxic air pollutants from formaldehyde.”⁴⁵⁷ After engaging in differential diagnosis to account for other possible causes, a medical doctor testified that it was more likely than not that the Wright family’s symptoms were caused by formaldehyde exposure.⁴⁵⁸ Requiring more than this demonstrates a misunderstanding of the probabilistic nature of scientific evidence.⁴⁵⁹

b. Absorption. The *Wright* court was unduly concerned about the difference between exposure to “pure” gaseous formaldehyde and formaldehyde attached to sawdust particles. Once formaldehyde gets into the lungs, however, it makes very little difference whether it got there in pure form or along with a sawdust carrier. The metabolism of the internal dose will be the same.⁴⁶⁰ The quantitative relationships between administered and delivered doses of formaldehyde are known.⁴⁶¹ Here the formaldehyde metabolites were found in the Wrights’ bodily fluids, showing that absorption did take place.

The effect of this decision is drastic for victims of environmental torts.⁴⁶² Information about exposure levels is commonly absent in

⁴⁵⁶ Plaintiffs clearly were exposed to enough of the emissions that particles of fiberboard laced with formaldehyde were found in the plaintiffs’ bodily fluids. See *Wright*, 91 F.3d at 1107 (noting that “wood fibers from defendant’s plant were in their house, their sputum, and their urine”); see also *id.* at 1109 (Heaney, J., dissenting) (observing that “significant levels of toxic emissions from the plant were found in their sputum and urine”). This is not a naturally occurring phenomenon. In addition, the testimony that formaldehyde was capable of causing the kinds of diseases from which the plaintiffs suffered was well supported. See *id.*

⁴⁵⁷ *Id.*

⁴⁵⁸ See *id.*

⁴⁵⁹ In addition, a threshold response was likely in light of the plaintiffs’ statements that their symptoms improved when they left their home and worsened when they returned. See Appellees’ Brief at 24, *Wright* (No. 95-4227). Such reversibility suggests traditional acute/chronic toxicity damage produced by processes reversible at low doses. See Rees & Hattis, *supra* note 279, at 277-78.

⁴⁶⁰ See discussion *supra* Part II.C.2.b.iv (discussing route of exposure).

⁴⁶¹ See generally Thomas B. Starr & Raymond D. Buck, *The Importance of Delivered Dose in Estimating Low-Dose Cancer Risk from Inhalation Exposure to Formaldehyde*, 4 *Fundamental Applied Toxicology* 740 (1984) (analyzing statistical relationship between delivered dose and administered dose of formaldehyde).

⁴⁶² For cases requiring evidence of exposure level based on *Wright*, see, e.g., *Mitchell v. Gencorp Inc.*, 165 F.3d 778, 781 (10th Cir. 1999) (requiring plaintiff to demonstrate “the levels of exposure that are hazardous to human beings generally as well as the plaintiff’s level of exposure to the defendant’s toxic substance before he or she may recover” (quoting *Wright*, 91 F.3d at 1106)); *Allen v. Pennsylvania Eng’g Corp.*, 102 F.3d 194, 199 (5th Cir. 1996) (“Scientific knowledge of the harmful level of exposure to a chemical, plus knowledge that the plaintiff was exposed to such quantities, are minimal facts necessary to sustain the plaintiffs’ burden in a toxic tort case.” (citing *Wright*, 91 F.3d at 1107)); *Mascarenas v. Miles, Inc.*, 986 F. Supp. 582, 587 (W.D. Mo. 1997) (supporting “obvious

toxic tort actions.⁴⁶³ In most cases, requiring that plaintiffs show the level of toxic chemical to which they were exposed is requiring them to achieve the impossible. Neither of these gaps would have been dispositive had the court used the default assumptions proposed here.

4. *Examine the Methodology*

None of the experts' methodologies were contested. The court did not examine or discuss the methodology of the underlying reports. It was willing to accept the published literature as validly demonstrating causation at least with respect to gaseous formaldehyde. Because formaldehyde has been so widely studied and the studies have been subjected to critique, there may be some legitimate basis for the court's failure to examine the methodology. Moreover, in this civil case, discovery was available to the parties, the opponents had access to each others' experts during depositions, and the experts disclosed the underlying studies upon which they relied. Because such access was available, one would have expected severe methodological flaws to come to light had there been any.

5. *Probabilistic Assessment of the Link Between Data and Hypothesis*

The plaintiffs bore the burden of establishing a valid basis for the experts' causation hypothesis that the defendant lumber mill's emissions caused the plaintiffs' illnesses. The data supporting the hypothesis include formaldehyde exposure by a plant emitting far greater amounts of formaldehyde than permissible and plaintiffs' symptoms that are consistent with formaldehyde exposure. Do any of the data refute the hypothesis that formaldehyde caused the plaintiffs' symptoms?⁴⁶⁴ Although the defendant argued that the symptoms were

proposition" that plaintiff in toxic substance case must prove exposure to that substance (citing *Wright*, 91 F.3d at 1106)); *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490, 1501-03 (E.D. Ark. 1996) (citing *Wright*, 91 F.3d at 1106), *aff'd per curiam*, 133 F.3d 1132 (8th Cir. 1998).

⁴⁶³ For cases excluding expert testimony where plaintiffs did not know their exposure level, see *supra* note 292 and accompanying text; see also *Wintz v. Northrop Corp.*, 110 F.3d 508, 512-14 (7th Cir. 1997) (affirming exclusion of expert's opinion as unreliable in part because expert had no information relating to plaintiff's specific exposure level); *Zwilling v. Garfield Slope Housing Corp.*, No. CV 94-4009, 1998 WL 623589, at *14 (E.D.N.Y. Aug. 17, 1998) (concluding expert's testimony was inadmissible due to plaintiff's unknown exposure level).

⁴⁶⁴ See Lanes, *supra* note 85, at 72 ("Empirical evidence about causal theories in medicine should be interpreted by describing testable, competing explanations."). Compare *Wright*, 91 F.3d at 1107-08, in which the expert testimony was excluded despite the absence of contrary data, with *Rutigliano v. Valley Bus. Forms*, 929 F. Supp. 779, 784-92 (D.N.J. 1996), *aff'd* 118 F.3d 1577 (3d Cir. 1997), in which the court painstakingly examined the contrary data before excluding the testimony. In *Rutigliano*, the plaintiff's

equally consistent with exposure to other naturally occurring agents, such as pollen, there was no data to support this.⁴⁶⁵ Additionally, a physician engaging in differential diagnosis disagreed with this pollen theory.⁴⁶⁶ The competing hypotheses should have gone—and did go—to the jury for its factfinding determination. Only on appeal did the court exclude this testimony.

For the general causation testimony to be admissible, several hypotheses had to be supported: first, that formaldehyde can cause human disease; second, that it can cause the kinds of diseases suffered by the plaintiffs; and third, that formaldehyde in the form and levels to which plaintiffs were exposed (formaldehyde carried on sawdust) can cause the kinds of illnesses the plaintiffs suffered. The first two the court acknowledged were established. The third prong was the issue the Eighth Circuit found the plaintiffs unable to meet,⁴⁶⁷ but had it understood the probabilistic reasoning underlying the concept of threshold responses, the court would have let the jury verdict stand.⁴⁶⁸

expert causation testimony was excluded because the expert failed to account for copious data that the plaintiff's symptoms predated her exposure to the formaldehyde in carbonless copy paper. It was not the plaintiff's inability to demonstrate the level of her exposure that was fatal to her case, but her inability to account for the data that carbonless copy paper was not implicated in the kind of injuries from which the plaintiff suffered. Plaintiff complained of "formaldehyde sensitization," which consists of symptoms of headache, tight throat, skin rash, fatigue, depression, and rhinitis, symptoms which, she claimed, made it impossible for her to work in an office, in a retail operation, or outdoors. See *id.* at 782-83. Moreover, Rutigliano's expert did not perform differential diagnosis on her patient and was unable to account for Rutigliano's allergy testing (which yielded negative results for formaldehyde sensitivity), her predominantly normal lung function tests, or why the plaintiff's Epstein Barr virus diagnosis did not account for her symptoms. See *id.* at 786-90. In addition, Rutigliano's expert failed to analyze her client's office environment, an analysis she usually undertook in detail. See *id.* at 791. Thus, because Rutigliano's expert's conclusions were untestable, and because her expert failed to account for data contrary to her hypothesis, data that indicated her conclusions might be incorrect, the court excluded the testimony. See *id.* The *Rutigliano* court's extensive consideration of all the data, including contrary data, stands in sharp contrast to *Wright*, in which the court cursorily dismissed plaintiffs' expert testimony even though no contrary data was presented, and without any extended analysis of the data at all.

⁴⁶⁵ The symptoms were, the manufacturer argued, common allergic symptoms. Thus, the argument was that something else—something unspecified—could have caused the plaintiffs' ailments. See Appellant's Opening Brief at 18, *Wright* (No. 95-4227).

⁴⁶⁶ See Appellees' Brief at 26-27, *Wright* (No. 95-4227).

⁴⁶⁷ Curiously, the Eighth Circuit acknowledged that it would not require a "mathematically precise table equating levels of exposure with levels of harm." *Wright*, 91 F.3d at 1107. However, it nevertheless insisted upon a showing of exposure level, a standard that few people exposed to environmental harms could meet.

⁴⁶⁸ The court neither discussed, nor does it appear to have applied, the traditional standard of review for the trial court's admissibility determination. Although this case was decided before *Joiner*, which reiterated the traditional abuse of discretion standard of review, it is nonetheless remarkable that the Eighth Circuit felt no compunction about its lack of deference to the trial judge. The Eighth Circuit instead reviewed the district court's

Specific causation here has yet another underlying hypothesis: that the formaldehyde-impregnated sawdust particles emanating from defendants' plant caused the plaintiffs' illness. In toxic tort cases, even if general causation is sufficiently established to gain acceptability as a theory (i.e., formaldehyde exposure causes disease in humans), direct evidence of specific causation is nearly always problematic.⁴⁶⁹ Plaintiffs nearly always demonstrate specific causation by the testimony of a physician who has engaged in differential diagnosis.⁴⁷⁰ The plaintiffs here indeed proffered such testimony. This testimony was supported by data that showed the presence of the lumber mill's product in the plaintiffs' home and bodies, a huge body of literature showing the effects of formaldehyde, symptoms consistent with formaldehyde exposure, and a temporal response that improved when the Wrights left the area.

The lumber mill defendants argued on appeal that because the plaintiffs' medical expert initially testified only that the plaintiffs' symptoms were "consistent" with exposure to formaldehyde, his testimony was irrelevant and therefore should have been excluded.⁴⁷¹ Defendants contended that "testimony as to 'consistency' (as opposed to causation)" was inadmissible.⁴⁷² Eventually—"after a great deal of prodding"—the plaintiffs' physician phrased the causation hypothesis in the terms the court demanded, stating that the symptoms were "more probably than not related to exposure to formaldehyde."⁴⁷³ Failing to understand the language of science and scientific conclusion-drawing, the Eighth Circuit concluded that the physician's testi-

"denial of a motion for judgment as a matter of law by applying the same standard that the district court applied originally." *Id.* at 1106. The highly deferential standard of review for admissibility determinations was subsequently reaffirmed by the Supreme Court in *Joiner*. See *General Elec. Co. v. Joiner*, 522 U.S. 136, 139 (1997). Had the *Wright* court applied the *Joiner* deference standard, it is unlikely that the Eighth Circuit would have reversed, because it would have had to find an abuse of discretion. Many courts employ the Eighth Circuit's analysis, however, in making initial *admissibility* determinations as well.

⁴⁶⁹ See, e.g., *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 715 (Tex. 1997) (acknowledging difficulties in adducing "direct, scientifically reliable" proof of specific causation in toxic tort cases).

⁴⁷⁰ See *Rutigliano*, 929 F. Supp. at 786 ("Courts have insisted time and time again that an expert may not give opinion testimony to a jury regarding specific causation if the expert has not engaged in the process of differential diagnosis—that is, the process of eliminating other possible diagnoses.").

⁴⁷¹ See Appellants' Opening Brief at 40, *Wright* (No. 95-4227).

⁴⁷² *Id.*

⁴⁷³ *Wright*, 91 F.3d at 1108.

mony was “speculation.”⁴⁷⁴ Had it understood the language of scientific argument, it would have concluded otherwise.⁴⁷⁵

Additionally, the Circuit Court set aside the jury verdict because the causation testimony involved gaseous formaldehyde rather than formaldehyde inhaled on wood particles.⁴⁷⁶ The court was unable to analogize from inhalation of gaseous formaldehyde to particulate inhalation, and failed to understand that the metabolism of the internal dose would not differ. Although this may be an extreme example, it underscores the general judicial unwillingness to reason by analogy in assessing scientific evidence.⁴⁷⁷ Because the *Wright* court had little understanding of the proffered evidence, and because it had no framework for analysis, it reached a decision that is scientifically—and from a policy standpoint—indefensible. Had it used the heuristic proposed in this Article, it could have done far better.

B. State v. Council

In this murder case, the court was faced with the question of whether a new technique for identifying hair by mtDNA analysis could meet standards of scientific validity.⁴⁷⁸ Hairs found at the crime

⁴⁷⁴ Id. The Eighth Circuit was apparently unfamiliar with the work of David Hume. Kenneth Rothman has noted that “Hume pointed out that observers cannot perceive causal connections, but only a series of events.” Kenneth J. Rothman, *Inferring Causal Inference—Habit, Faith or Logic?*, in *Causal Inference*, supra note 75, at 3, 5.

⁴⁷⁵ The court failed to understand that an intellectually honest scientist cannot testify to causation. See supra notes 377-84 and accompanying text.

⁴⁷⁶ See *Wright*, 91 F.3d at 1107-08.

⁴⁷⁷ For other cases exhibiting this failure, see *Sutera v. Perrier Group of Am. Inc.*, 936 F. Supp. 655, 664 (D. Mass. 1997) (failing to reason by analogy from animal studies to plaintiff's exposure); *Mascarenas v. Miles, Inc.*, 986 F. Supp. 582, 587 (W.D. Mo. 1997); *Mancuso v. Consolidated Edison Co.*, 967 F. Supp. 1437, 1445 (S.D.N.Y. 1997), aff'd in part and vacated in part on other grounds, 216 F.3d 1072 (2d. Cir. 2000). The Third Circuit, in contrast, clearly understands the importance of scientific analogy. In *re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717 (3d Cir. 1994), involved plaintiffs exposed to PCBs. As Judge Becker explained for the court in *Paoli*, the judge not only must assess the investigative process used by the expert, but also must make an independent assessment of the reliability of the expert's data and understand which inferences justifiably may be drawn from it. See *Paoli*, 35 F.3d at 747-48.

⁴⁷⁸ See *State v. Council*, 515 S.E.2d 508, 516-19 (S.C.), cert. denied, 120 S. Ct. 583, 588-89 (1999). This was the second instance in which the FBI laboratories had proffered such evidence at a criminal trial. See id. at 516 & n.13 (citing Mark Curriden, *A New Evidence Tool: First Use of Mitochondrial DNA in a U.S. Criminal Trial*, A.B.A. J., Nov. 1996, at 18 (reporting first such trial)); see also supra note 404. The court emphasized that although South Carolina utilizes neither the *Frye* nor the *Daubert* standard, see id. at 517-18, it does make admissibility determinations based on a South Carolina rule “identical to Rule 702 of the Federal Rules of Evidence,” id. at 517. Although not adopting *Daubert*, the South Carolina standard for admitting scientific evidence is whether “the evidence will assist the trier of fact, the expert witness is qualified, and the underlying science is reliable.” Id. at 518. Factors in this determination include “(1) the publications and peer review of the technique; (2) prior application of the method to the type of evidence involved in the case;

scene were subjected to both conventional hair analysis and to mtDNA testing and compared with samples of the defendant's hair.⁴⁷⁹ Mitochondrial DNA analysis, new to the courts, offers several advantages over conventional DNA analysis.⁴⁸⁰ As the *Council* court noted, because there is only one nucleus per cell but many mitochondria, mtDNA can be obtained in much larger quantities.⁴⁸¹ Moreover, mtDNA is more stable than nuclear DNA, and while nuclear DNA (of hair) is only found in the living cells at the base of a hair follicle, mtDNA can be found in the hair shaft.⁴⁸²

The court admitted the evidence over objection, despite the defendant's inability to procure an expert in his defense, and despite a very curtailed opportunity to review the test data.⁴⁸³ The court found the testimony admissible not on any independent validity analysis, but based on the defendant's opportunity to attack any "'shaky but admissible evidence'" on cross-examination.⁴⁸⁴ The prosecution expert testified that "most probably" the hair recovered from the crime scene was the defendant's.⁴⁸⁵ Although the prosecution expert testified that, of those hairs that could be sequenced, the reliability of getting a correct sequence was 100%,⁴⁸⁶ he did admit on cross-examination that it was possible that the hair belonged to another individual.⁴⁸⁷

(3) the quality control procedures used to ensure reliability; and (4) the consistency of the method with recognized scientific laws and procedures." *Id.* at 517 (citing *State v. Ford*, 392 S.E.2d 781, 783 (S.C. 1990)). Thus, it is essentially the same as the *Daubert* standard.

⁴⁷⁹ See *id.* at 516.

⁴⁸⁰ Mitochondrial DNA differs from the nuclear DNA conventionally used in suspect identification in that mtDNA comes from the mitochondria of the cell rather than from the cell nucleus. See *id.* at 516 n.12 (citing Brian Huseman, *Taylor v. State*, Rule 706, and the DNA Database: Future Directions in DNA Evidence, 22 Okla. City U. L. Rev. 397, 409 (1997) (noting potential utility of new developments in DNA sequencing)).

⁴⁸¹ See *id.*

⁴⁸² See *id.*

⁴⁸³ See *id.* at 518. The mtDNA analysis was not released to the defense until the night before trial, leaving the defense no opportunity to seek an expert. See *id.* This highlights one of the major shortcomings of the use of scientific testimony in criminal trials: the inability of the defense to procure an expert. See Paul C. Giannelli, *The Abuse of Scientific Evidence in Criminal Cases: The Need for Independent Crime Laboratories*, 4 Va. J. Soc. Pol'y & L. 439, 473-74 (1997) (proposing independent labs, available to defense, to remedy abuses of scientific evidence by prosecution experts). Even after the Supreme Court recognized the right of the defense to some expert assistance under certain circumstances in *Ake v. Oklahoma*, 470 U.S. 68, 74 (1985), indigent defendants' access to scientific experts is frequently severely limited, as it was in *Council*. See Paul C. Giannelli, *The DNA Story: An Alternative View*, 88 J. Crim. L. & Criminology 380, 414-21 (1997) (book review) (noting difficulties defendants often face in obtaining information about prosecution expert test data and in accessing experts of their own).

⁴⁸⁴ See *Council*, 515 S.E.2d at 519 (quoting *State v. Dinkins*, 462 S.E.2d 59, 60 (S.C. 1995)).

⁴⁸⁵ See *id.* at 517.

⁴⁸⁶ See *id.*

⁴⁸⁷ See *id.* at 519.

1. *What's the Theory?*

Hair identification rests upon the theory that each individual has unique characteristics, but this is an assumption that is largely untested.⁴⁸⁸ The prosecution proffered two types of hair analysis, microscopic and mtDNA.⁴⁸⁹ With respect to the microscopic hair analysis, the prosecution theory was that the microscopic characteristics of hair differ among people. An FBI "white paper commentary" on microscopic hair analysis, however, concedes that although the "microscopic characteristics of hair can assist the examiner in determining the racial origin, body area and whether disease, damage or artificial treatment is present," they cannot be used to identify an individual.⁴⁹⁰ Although these characteristics may be "useful," the FBI recognizes that "hair comparisons do *not* constitute a basis for absolute personal identification."⁴⁹¹ As a result, microscopic hair comparisons no longer can be used alone and now must be used by the FBI laboratory in conjunction with mtDNA analysis.⁴⁹²

The second part of the prosecution's identification hypothesis rested on the theory that each person has unique genetic characteristics that can be determined by sequencing the mtDNA in a hair follicle. The theory behind DNA analysis in general is that genetic differences exist between people and that DNA analysis can uncover those differences.⁴⁹³ Unlike microscopic hair analysis, this theory is

⁴⁸⁸ See *supra* notes 336-47 and accompanying text.

⁴⁸⁹ See *Council*, 515 S.E.2d at 516.

⁴⁹⁰ Douglas Deedrick, FBI Responds to Questions Raised About Hair Comparison Analysis, Prosecutor, Nov.-Dec. 1998, at 27, 27.

⁴⁹¹ *Id.* (emphasis added).

⁴⁹² See *id.*

⁴⁹³ For an introduction to the forensic use of DNA evidence, see generally William C. Thompson, Guide to Forensic DNA Evidence, in *Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual 195* (Bert Black & Patrick W. Lee eds., 1997). In brief, the human genome is made up of approximately four billion organic base pairs in a particular sequence. The base pairs all consist of adenine (A), thymine (T), cytosine (C), or guanine (G), A pairing with T, and C with G on the two complementary strands of DNA. See C.G.G. Aitken, *Statistics and the Evaluation of Evidence for Forensic Scientists* 10 (1995). Most of the base-pair sequence is common to all human beings. See Thompson, *supra*, at 203. The "genome" is the full complement of human DNA. See *id.* The mitochondrial genome is made up of some 16,000 base pairs. See *id.*

At some locations on the genome, there are distinctive sequences of base pairs, known as "alleles." See *id.* at 203. Forensic DNA testing examines these alleles. In order to estimate the relative frequency of a particular genotype by counting its occurrence in the population, a huge database of samples would be needed. See George Sensabaugh & David H. Kaye, Non-Human DNA Evidence, 39 *Jurimetrics* 1, 11-12 (1998) (explaining statistical basis for DNA identification techniques). Therefore, instead of the complete genotype, alleles—a grouping of distinct DNA characteristics at particular locations on the genome—are ordinarily used. See *id.* Typically, two alleles are found at each location, one inherited from the mother and one from the father. By comparing the groupings of the alleles of two samples at a number of different locations, the analyst can determine

well accepted.⁴⁹⁴

Mitochondrial DNA analysis is different from nuclear DNA analysis, however, because in mtDNA all the DNA comes from the mitochondria of the cell rather than the cell nucleus. Mitochondria are cellular structures with their own DNA.⁴⁹⁵ All the DNA in mtDNA analysis comes from the mother, whereas in nuclear DNA analysis, half comes from each parent.⁴⁹⁶ Therefore, all offspring from the same mother have the same mtDNA, unless there have been genetic mutations.⁴⁹⁷ Thus, unlike nuclear DNA, mtDNA varies only moderately among different individuals, with some areas of the mitochondrial genome being more variable than others.⁴⁹⁸

2. *Examine All the Available Evidence*

Samples of the suspect's hair in *Council* were analyzed and compared with samples of hair taken from the crime scene.⁴⁹⁹ Microscopic hair analysis was offered in conjunction with the mtDNA analysis, and the expert concluded that the defendant's hair was the same as that found at the crime scene.⁵⁰⁰ The expert explained that, for the mtDNA analysis, he (or someone else at his lab) extracted DNA from the mitochondria of the hair cells, amplified it, and ex-

whether or not the two samples are from the same source. This is referred to as a "match," but what it really means is that the null hypothesis cannot be substantiated. Then the analyst performs statistical analysis to determine the probability that the nonexclusion would occur by chance in the population. See *id.* at 11. This frequency is estimated through the use of a database. Combining the allele frequencies into the final profile frequency (the random match probability) is a mathematical function of the genetic diversity at each location on the genome and the number of the locations tested. See Committee on DNA Forensic Science, National Research Council, *The Evaluation of Forensic DNA Evidence* 146-47 (1996). The method used by the expert in *Council* differed in that the sequences of base pairs in the defendant's and the crime scene's samples were directly compared with each other and the population frequency of such sequences assessed through the database.

⁴⁹⁴ See Aitken, *supra* note 493, at 11 (discussing acceptance of "evidential value" of DNA profiling).

⁴⁹⁵ See Charles M. Strom, *Genetic Justice: A Lawyer's Guide to the Science of DNA Testing*, Ill. B.J., Jan. 1999, at 18, 24.

⁴⁹⁶ The *Council* court recognized this difference. See *State v. Council*, 515 S.E.2d 508, 516 n.12 (S.C.), cert. denied, 120 S. Ct. 588 (1999).

⁴⁹⁷ See Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 Mich. L. Rev. 931, 934 n.16 (1999) (noting potential utility of mtDNA testing). This is a crucial—and controversial—qualification. At least two studies have found significantly higher mutation rates in mtDNA than in nuclear DNA. See Ann Gibbons, *Calibrating the Mitochondrial Clock (DNA Mutations and Evolutionary Dating)*, 279 Science 28, 29 (1998) (outlining recent research suggesting rapid mutation rates in mtDNA).

⁴⁹⁸ See Strom, *supra* note 495, at 24 (laying out basic information on DNA identification).

⁴⁹⁹ See *Council*, 515 S.E.2d at 516.

⁵⁰⁰ See *id.* at 516, 517.

amined it to determine its allele sequence.⁵⁰¹ The sequence was compared to the sequence of a sample obtained from the defendant, and then the sequence was compared to the database of known mtDNA sequences, which contained 742 known sequences.⁵⁰² Of the hairs that could be sequenced, the expert testified that the “reliability of getting a correct sequence was 100%.”⁵⁰³ Three hundred and nineteen of the sequences were from African-Americans, and the defendant was an African-American.⁵⁰⁴ The expert had found matching sequences between unrelated Caucasians in the database previously, but had never found a match between unrelated African-Americans.⁵⁰⁵ Only two regions were analyzed, and, according to the expert, they were the “most variable [regions] in African-Americans.”⁵⁰⁶

The prosecution expert testified that mtDNA analysis has been used for research since 1981, and that it currently is used to identify bodies from wars.⁵⁰⁷ At the time of the expert’s testimony, over 600 papers had been published on mtDNA, although it was not clear from the court’s account of the testimony whether these papers focused on identification, evolution, or some other use of the technique.⁵⁰⁸ Presumably, these studies formed the basis of the expert’s testimony, although the opinion does not say so. Mitochondrial DNA analysis has been used to identify the remains of people massacred in human rights violations in the Balkans⁵⁰⁹ and Latin America,⁵¹⁰ as well as in identifying the bodies of Tsar Nicholas II of Russia and his family.⁵¹¹ It is also widely used for inferring evolutionary relationships among species and populations because the sequence of mtDNA alleles changes only through mutation, and not through the genetic mixing of sexual

⁵⁰¹ See *id.* at 517. For background on alleles and DNA testing, see *supra* note 493.

⁵⁰² See *id.*

⁵⁰³ *Id.* What can this statement mean? Presumably, a judge with an understanding of statistics and probability theory would have to view this statement with skepticism, as no test can be 100% reliable. See *supra* notes 77-87, 197-204 and accompanying text.

⁵⁰⁴ See *Council*, 515 S.E.2d at 517 & n.14.

⁵⁰⁵ See *id.* at 517.

⁵⁰⁶ *Id.*

⁵⁰⁷ See *id.* at 516.

⁵⁰⁸ See *id.*

⁵⁰⁹ See Eliot Marshall, Tracing Croatia’s “Disappeared,” 269 *Science* 1812, 1812 (1995) (suggesting mtDNA may be more useful than nuclear DNA in identifying bodies from mass graves).

⁵¹⁰ See Ann Gibbons, Scientists Search for “The Disappeared” in Guatemala, 257 *Science* 479, 479 (1992) (using mtDNA to identify victims of military killings buried in unmarked graves).

⁵¹¹ See Gibbons, *supra* note 497, at 28 (describing use of mtDNA testing to identify soldiers’ remains and remains of nine Russians exhumed from Siberian grave thought to be of Tsar Nicholas II and his family).

reproduction from generation to generation.⁵¹² Scientists have examined animal hairs for genetic information for more than a decade.⁵¹³ Thus, there appears to be some support for the use of this technique in identifying individuals.

3. *Use Valid Assumptions to Fill the Gaps*

The individualization assumption of the prosecution expert, that mtDNA testing can be used to identify the source of crime scene hairs, appears warranted. This assumption rests on the theory that each person's complete genome is unique, and that although there are many common areas, some locations on the genome are more variable than others.⁵¹⁴ Mitochondrial DNA is not quite as strong an identifying technique as nuclear DNA, because all offspring of the same mother will share the same sequences of mtDNA, barring mutations.⁵¹⁵ Although the assumption that DNA analysis is capable of identifying individual characteristics has strong empirical support, methodological flaws may undermine its application severely.

4. *Examine the Methodology*

The technique used to analyze mtDNA is known as polymerase chain reaction (PCR), which takes a small amount of DNA and amplifies it in a test tube.⁵¹⁶ This process makes it extraordinarily sensitive to contamination.⁵¹⁷ The amplified DNA is then examined to deter-

⁵¹² See, e.g., Patricia Kahn & Ann Gibbons, DNA from an Extinct Human, 277 *Science* 176, 176 (1997) (reporting recent research using ancient mtDNA to suggest Neanderthals are not direct ancestors of modern humans).

⁵¹³ See Russell Higuchi et al., DNA Typing from Single Hairs, 332 *Nature* 543, 545 (1988) (noting that hair offers many advantages, because it is more easily found, transported, and stored than blood).

⁵¹⁴ See David H. Kaye, DNA, NAS, NRC, DAB, RFLP, PCR, and More: An Introduction to the Symposium on the 1996 NRC Report on Forensic Evidence, 37 *Jurimetrics* 395, 399 n.36 (1997) (explaining that expected frequency of DNA profile at several locations on chromosome is product of frequencies of different alleles).

⁵¹⁵ See Gibbons, *supra* note 497, at 28 ("[M]tDNA comes only from the mother's egg."); Mark Hansen, A Comeback for Hair Evidence, *A.B.A. J.*, May 1998, at 66, 67 ("Nuclear DNA testing is more precise than mitochondrial testing, experts say, because the DNA found in the nucleus . . . is unique to every individual except identical twins . . . [while] mitochondrial DNA . . . is the same for all maternal relatives . . ."); cf. Richard Lempert, After the DNA Wars: Skirmishing with NRC II, 37 *Jurimetrics* 439, 455-58 (1997) (explaining that because nuclear DNA is inherited from both parents, few unrelated individuals are expected to share same sequences at highly variable locations).

⁵¹⁶ See Thompson, *supra* note 493, at 215-16 (explaining that PCR process that is used in mtDNA analysis involves breaking double-stranded DNA fragments into single strands and then inducing each single strand to bind with complementary base pairs floating in solution).

⁵¹⁷ See Committee on DNA Tech. in Forensic Science, National Research Council, DNA Technology in Forensic Science 65-67 (listing three types of contamination and suggesting ways to avoid it).

mine the sequence of base pairs. Proper procedures are critical because contamination from other biological material through handling or from previously performed amplifications will cause the contaminating DNA to be amplified along with the sample.⁵¹⁸ Contamination in mtDNA testing increases the chances that sequences will be incorrect and may increase the chance of declaring a false match. Ordinarily, controls are run with the test sample to ensure that contamination does not occur.

The FBI expert in *Council* testified that DNA extracted from the mitochondria in the crime-scene hair was extracted, amplified, and examined to determine its base-pair sequences.⁵¹⁹ He made no mention of controls. Two samples, one from the suspect and the other from the murder scene, were compared.⁵²⁰ The expert did not discuss laboratory protocols, standards, or proficiency testing. The examiner then compared the sequence to 742 known sequences in the database and concluded that the two sequences matched.⁵²¹ The expert did not discuss the frequency of the examined sequences in the population. No error rate was proffered, although the court asserted that the FBI laboratory had "validated the process and determined its [own] rate of error."⁵²²

The National Research Council has issued two reports on DNA evidence calling for the utilization of high quality laboratory standards, proficiency testing, and sample splitting for independent testing.⁵²³ The court in *Council* did not question the expert on these issues or examine any of the expert's laboratory notes before ruling on admissibility, nor was this information available to the defense. Rather, the court relied on cross-examination to raise any problems.⁵²⁴ Releasing the test results the night before trial, however, effectively denied the defendant meaningful access to the expert's methodology and precluded any detailed cross-examination by the defendant.⁵²⁵ The test results alone, without the information about pro-

⁵¹⁸ See *id.* at 65-66.

⁵¹⁹ See *State v. Council*, 515 S.E.2d 508, 517 (S.C.), cert. denied, 120 S. Ct. 588 (1999).

⁵²⁰ See *id.* A third sample from another suspect also was compared, but the results excluded the other suspect as a possibility because there was no match. See *id.*

⁵²¹ See *id.* Note that the expert referred to "known sequences," rather than the number of people in the database. See *id.* This may indicate a problem with database size. See *supra* notes 343-44 and accompanying text.

⁵²² *Id.* at 518.

⁵²³ See Committee on DNA Forensic Science, *supra* note 493, at 4; see also Erika Hagelberg et al., Identification of the Skeletal Remains of a Murder Victim by DNA Analysis, 352 *Nature* 427, 427 (1991) (discussing contamination of PCR methods).

⁵²⁴ See *Council*, 515 S.E.2d at 519.

⁵²⁵ See Paul C. Giannelli, Criminal Discovery, Scientific Evidence, and DNA, 44 *Vand. L. Rev.* 791, 803-04 (1991) (arguing that mere notice that expert will testify for prosecution

cedures and controls, were simply not enough to reveal the potential weaknesses of the testimony.⁵²⁶ For example, control samples without added DNA normally are processed at several stages in the testing process, and the entire test may be discarded if DNA is found in the controls.⁵²⁷

Not only were the absence of controls and protocols suspect, but the expert's characterization of his methodology vaunted its accuracy to an unlikely degree when he testified that "the reliability of getting a correct sequence was 100%."⁵²⁸ It is unlikely that any procedure is 100% accurate, and the expert neither offered support for his statement, nor was any demanded by the court. Without further explanation and access to the methodology, this testimony is inherently untrustworthy.⁵²⁹

5. *Probabilistic Assessment of the Link Between Data and Hypothesis*

The failure to grant a defense expert access to the underlying laboratory notes critically weakens the argument for admissibility here. Failure to follow protocols threatens the validity of DNA analysis.⁵³⁰ Because the mtDNA technique's weakest aspect is its sensitivity to contamination, it would be crucial to know whether the testing observed adequate controls. No meaningful results can be drawn from a test run without adequate controls.⁵³¹ Moreover, proficiency testing—a form of outside audit that tests the technician's ability to perform the procedure correctly—was apparently not done. The court never sought any information on controls or methodology.

The expert in *Council* concluded that "most probably" the hair recovered from the crime scene was that of the defendant.⁵³² The

is insufficient and that more access to laboratory results is needed than is provided in typical laboratory report, because scientific reports rarely indicate details of methodology).

⁵²⁶ Defendant's council argued—correctly, in my opinion—that the admissibility determination could not be made without allowing the defense to present rebuttal evidence, including an expert witness. See *Council*, 515 S.E.2d at 518 n.16, 519.

⁵²⁷ See Thompson, *supra* note 493, at 221-22 & 222 n.51 (noting that when negative controls "fail" by detecting the presence of DNA, "contamination is strongly suggested and conclusion-drawing in such situation is "extremely dangerous").

⁵²⁸ *Council*, 515 S.E.2d at 517. Presumably the expert meant that if a hair included a particular allele sequence, such as ATA, this sequence always would appear as ATA in the test results.

⁵²⁹ See Giannelli, *supra* note 525, at 804.

⁵³⁰ See, e.g., *United States v. Martinez*, 3 F.3d 1191, 1197-98 (8th Cir. 1993) (acknowledging that courts must determine whether expert failed to follow protocols and, if so, whether error made results unreliable).

⁵³¹ See Thompson, *supra* note 493, at 252 (stating that at least three types of controls—positive, negative, and substrate—are necessary to ensure test's reliability).

⁵³² See *Council*, 515 S.E.2d at 517.

court admitted this testimony despite the expert's concession that he had found matches in other mtDNA analyses between unrelated people.⁵³³ Although nuclear DNA identification techniques are well supported, and mtDNA methodology has been applied successfully in contexts other than litigation,⁵³⁴ admissibility of this particular mtDNA testimony was ill advised. Mitochondrial DNA testing is not as powerful a tool as nuclear DNA testing because people with no maternal link still may have significant numbers of base pair matches,⁵³⁵ and the prosecution's expert conceded as much.

Moreover, the expert's assertion that mtDNA testing is extremely reliable is correct only if the mtDNA test is correctly performed and interpreted. The PCR method used in mtDNA testing is extraordinarily difficult to perform without error from contamination.⁵³⁶ The error rate of false positive "matches" is a serious concern in DNA evidence.⁵³⁷

Another problem in *Council* was the size of the database used for hair comparison, which consisted of only 742 known sequences. Larger samples give more precise estimates of allele frequencies than small ones.⁵³⁸ Moreover, a proper method of constructing a database

⁵³³ See *id.* at 518. The expert qualified this statement, however, by saying that he never had found matches between unrelated African-Americans. Because the defendant was African-American, this statement was highly misleading. It implies that there may be more similarities between the mtDNA of Caucasians than between that of African-Americans, a point that never has been demonstrated. The more likely cause was a small database of African-Americans; the database used by the expert in *Council* consisted of only 319 sequences obtained from African-Americans. See *id.*; cf. Gibbons, *supra* note 497, at 29 (noting that observed high incidence of mtDNA mutations in studies of families may be "statistical artifact" that will disappear as studies increase in size).

⁵³⁴ See *supra* note 509-11 and accompanying text.

⁵³⁵ See Jon Cohen, Genes and Behavior Make an Appearance in the O.J. Trial, 268 *Science* 22, 22 (1995) (observing that there have been cases of matching mtDNA for more than 200 base pairs in absence of maternal link, and that "most frequent [mtDNA] sequence . . . appears in 3% of the population").

⁵³⁶ See Kahn & Gibbons, *supra* note 512, at 177 (noting difficulty in avoiding errors in mtDNA testing of ancient Neanderthals); see also Giannelli, *supra* note 525, at 796-97 (noting false identification problems in proficiency tests).

⁵³⁷ See Jonathan J. Koehler, Error and Exaggeration in the Presentation of DNA Evidence at Trial, 34 *Jurimetrics* 21, 26 (1993) (estimating rate of false positive matches); William C. Thompson, Accepting Lower Standards: The National Research Council's Second Report on Forensic DNA Evidence, 37 *Jurimetrics* 405, 417 (1997) (noting that "the probability of an erroneous match is difficult to estimate"). The expert really should have addressed how likely it was that the sequence would be reported the same in the two hair samples if the mtDNA were not the defendant's. See Lempert, *supra* note 515, at 442 ("A scientist who testifies that false positive error never happens does not address the question the jury needs answered—namely, how likely is it that a match would be reported if the evidence DNA was not the suspect's.")

⁵³⁸ See *supra* Part II.C.1; cf. Committee on DNA Forensic Science, *supra* note 493, at 114 (observing need for confidence intervals with respect to estimates if database is small). But see Ranajit Chakraborty, Sample Size Requirements for Addressing the Population

would be to construct a “local database of DNA samples with representative genotype frequencies from the local population” or the geographical area of the crime.⁵³⁹ Yet the prosecution offered no information about how the database was constructed. In addition, no explanation was advanced—at least none was noted by the court—for the extent of statistical error.

Instead, the court accepted at face value the expert’s assertion that the “FBI laboratory validated the process and determined its rate of error” without any inquiry into whether proficiency testing had been done.⁵⁴⁰ Proficiency testing is important because it demonstrates that a technique is valid not only in theory, but in practice.⁵⁴¹ In sum, *Council* is an instance where the court’s failure to conduct a meaningful inquiry into expert methodology may have significantly undermined its gatekeeping duties.⁵⁴²

CONCLUSION

The Supreme Court’s evidentiary trilogy—*Daubert*, *Joiner*, and *Kumho Tire*—has made an important contribution toward rationalizing the jurisprudence of scientific evidence. The Supreme Court has demonstrated emphatically that neither judges nor lawyers can afford to be ignorant of the scientific process.⁵⁴³ These cases must be read as exhortatory rather than illuminating, however, as they offer little in

Genetic Issues of Forensic Use of DNA Typing, 64 *Hum. Biology* 141, 156 (1992) (suggesting that relatively small databases may allow statistically acceptable frequency estimation for common alleles).

⁵³⁹ Strom, *supra* note 495, at 20 (holding that such database could determine if identification of local resident was erroneous match).

⁵⁴⁰ *State v. Council*, 515 S.E 2d 508, 518 (S.C.), cert. denied, 120 S. Ct. 588 (1999). The National Research Council explained that “there is no substitute for rigorous proficiency testing, via blind trials.” Committee on DNA Tech. in Forensic Science, *supra* note 517, at 55. Proficiency testing is similar to the double-blind trials discussed earlier. See *supra* note 219 and accompanying text.

⁵⁴¹ See Committee on DNA Tech. in Forensic Science, *supra* note 517, at 55 (noting that “proficiency testing constitutes scientific confirmation that a laboratory’s implementation of a method is valid not only in theory, but also in practice”).

⁵⁴² Mitochondrial DNA now has been found scientifically reliable in state courts in Maryland, Michigan, New York, North Carolina, Pennsylvania, and Tennessee, as well as in the U.S. District Court for the Eastern District of Ohio. See Leigh Jones, *Type of DNA Ruled Reliable in Rape Trial*, N.Y. L.J., Sept. 7, 2000, at 1. Yet the methodological problems with the mtDNA evidence proffered in *Council* should have led the court to exclude the evidence.

⁵⁴³ On the importance of scientific knowledge to judges and lawyers, see Leonard R. Jaffee, *Of Probativity and Probability: Statistics, Scientific Evidence, and the Calculus of Chance at Trial*, 46 *U. Pitt. L. Rev.* 925, 928, 929 (1985) (“[M]ath is not a magic [judges and lawyers] must bow to as cavemen to thunder—even that their logic, intuition and experience make better courtroom medicine than can a number-juggler’s calculus. . . . [N]o calculus of probability is competent to measure the sufficiency or weight of litigational evidence.”).

the way of guidance for performing the analysis that they require. The guidelines proposed in this Article attempt to offer a more usable framework for such analysis.

The heuristic proposed here advances a method of assessing the strength of reasoning by analogy that is the hallmark of scientific argument. Rather than attempting to explain what constitutes the sound design of good experiments, this framework accepts as a given that experiments as performed never measure up to the guidelines in all respects, that information is never complete, and that analogies are always imperfect. This framework offers judges, and the lawyers who must inform them, a way of dealing with the imperfections, making it possible for judges to evaluate the evidence that is available.

Because scientific knowledge is never complete, judges inevitably make assumptions to fill the gaps. Many of the assumptions judges currently use are unstated and make little scientific or policy sense. To treat litigants more fairly, to give them the intellectual process they are due, courts need to make their unstated assumptions explicit. Moreover, judges need to adopt and rely on default assumptions that are firmly based on good science and good policy. This heuristic allows judges and lawyers to examine imperfections, to discard unjustifiable inferences, and to weigh the cumulative force of justifiable inferences. Such a process is the way scientists critique each others' work. Nonscientists who must make scientific validity determinations also can use this process to illuminate their decisions. In short, this heuristic offers judges a way to give intellectual due process to competing explanations of reality.