

GATEKEEPING SCIENCE: USING THE STRUCTURE OF SCIENTIFIC RESEARCH TO DISTINGUISH BETWEEN ADMISSIBILITY AND WEIGHT IN EXPERT TESTIMONY

David L. Faigman, Christopher Slobogin & John Monahan

ABSTRACT—Fundamental to all evidence rules is the division of responsibility between the judge, who determines the admissibility of evidence, and the jury, which gauges its weight. In most evidentiary contexts, such as those involving hearsay and character, threshold admissibility obligations are clear and relatively uncontroversial. The same is not true for scientific evidence. The complex nature of scientific inference, and in particular the challenges of reasoning from group data to individual cases, has bedeviled courts. As a result, courts vary considerably on how they define the judge’s gatekeeping task under Federal Rule of Evidence 702 and its state equivalents.

This Article seeks to reconceptualize gatekeeping analysis in scientific evidence cases based on the nature of science itself, specifically, the division between general and case-specific scientific findings. Because expert testimony describing basic science, “framework” science, and the scientific methods an expert uses to reach conclusions transcend the case at hand, the validity of these facts ought to be determined by the judge. In contrast, when an expert claims to have used a methodology approved by the judge, but a dispute arises as to whether the expert in fact did so, the question becomes one of credibility specific to the case, and is for the jury.

This division between general and case-specific preliminary facts is simpler to administer than other admissibility–weight frameworks, which have relied primarily on problematic attempts to distinguish scientific methods from scientific conclusions. It is also fully consistent with, and helps implement, basic principles of both constitutional and evidentiary jurisprudence by ensuring that the trial judge—presumptively better attuned to matters of general import—decides reliability issues, while the jury—historically viewed as trier of the facts—is the ultimate arbiter of those case-specific matters requiring a credibility assessment. Because the general–specific divide likewise argues for a stiff standard of appellate review on scientific reliability issues, our alignment of evidence law with the nature of scientific research also provides the best court-monitored

mechanism for ensuring that courtroom use of science is both sophisticated and consistent across cases.

AUTHORS—David L. Faigman is the John F. Digardi Distinguished Professor of Law, University of California Hastings College of the Law; Professor, University of California San Francisco, School of Medicine, Department of Psychiatry. Christopher Slobogin is the Milton R. Underwood Professor of Law, Affiliate Professor of Psychiatry, Vanderbilt University. John Monahan is the John S. Shannon Distinguished Professor of Law, Professor of Psychology, and Professor of Psychiatry and Neurobehavioral Sciences, University of Virginia. We thank Edward Cheng, Lisa Faigman, Roger Park, Gregory Mitchell, and members of the MacArthur Neuroscience Network for their comments on earlier drafts of this Article. Preparation of this Article was supported, in part, by a grant from the John D. and Catherine T. MacArthur Foundation to Vanderbilt University. Its contents reflect the views of the authors, and do not necessarily represent the official views of either the John D. and Catherine T. MacArthur Foundation or the MacArthur Foundation Research Network on Law and Neuroscience (www.lawneuro.org).

INTRODUCTION	861
I. THE GENESIS OF THE METHODOLOGY–CONCLUSIONS DISTINCTION	867
II. PREMISES THAT INFORM THE ADMISSIBILITY OF SCIENTIFIC EVIDENCE.....	874
A. <i>Constitutional Considerations</i>	875
B. <i>Evidentiary Considerations</i>	880
III. ALIGNING THE RULES OF EVIDENCE WITH SCIENTIFIC INFERENCE.....	885
A. <i>The Structure of Scientific Evidence</i>	885
B. <i>Categories of Science</i>	889
C. <i>A Test Based On Scientific Inference</i>	900
IV. IMPLICATIONS FOR APPELLATE REVIEW STANDARDS	901
CONCLUSION	903

INTRODUCTION

A mainstay of the law of evidence is the distinction between admissibility and weight. Judges are tasked with the responsibility of determining whether proffered evidence is *admissible* and, if it is, jurors must decide what *weight* to give it.¹ This division of responsibility holds across all evidence rules, from basic relevance to hearsay. Hence, a “dying declaration” that is hearsay is only admissible if the judge determines, among other things, that the statement was made by a declarant “while believing that the declarant’s death was imminent.”² If admitted, the weight, if any, that should be accorded the particular dying declaration is up to the jury to decide. The predicate issue of whether the defendant believed death was imminent is called a “preliminary fact,”³ one that the judge must determine by a preponderance of the evidence.⁴ This

¹ Throughout this Article, we refer to jurors and juries as the factfinder at trial, since the admissibility–weight issue is virtually never litigated unless that is the case. However, the analysis offered here is meant to apply whether the trier of fact is a judge or jury. Even though the judicial and verdict roles merge in a bench trial, judges should not consider inadmissible scientific evidence any more than juries should. Furthermore, under our analysis in Part IV, the deference the rulings of a judge at a bench trial would receive at the appellate level would vary depending on whether the ruling deals with admissibility or weight.

² FED. R. EVID. 804(b)(2).

³ According to Black’s Law Dictionary, a preliminary or predicate fact is “[a] fact necessary to the operation of an evidentiary rule.” *Predicate Fact*, BLACK’S LAW DICTIONARY (10th ed. 2015).

⁴ This is the standard in federal court at least. It should also be noted that the preliminary fact standard is applied differently in different contexts. When the evidence rule deals with expert testimony or hearsay, the judge must find the preliminary fact by a preponderance of the evidence. *See Bourjaily v. United States*, 483 U.S. 171, 175–76 (1987) (holding that the preponderance of the evidence standard should govern hearsay admissibility questions under Rule 104). However, when the rule deals with

prerogative belongs to the judge on the assumption that juries are likely to attribute significance to an out-of-court statement even if they were to find that death was not imminent.⁵

This division of responsibilities between judge and jury also applies to scientific evidence presented by experts.⁶ Indeed, because the distinction between admissibility and weight is endemic to the law of evidence, in theory the issue has existed for scientific evidence since experts were first introduced at trial in the nineteenth century.⁷ Yet the distinction received little attention until the United States Supreme Court's landmark decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*⁸ The *Daubert* Court, interpreting Federal Rule of Evidence 702, held that judges are “gatekeepers” and obligated to determine whether the methods and principles underlying proffered expert testimony are—more likely than not—reliable and valid.⁹ In other words, the Court treated the “evidentiary reliability” of the scientific evidence proffered in the case as a preliminary fact and thus within the judge's purview to determine. Consistent with the preliminary fact rule in other types of cases, the rationale for this requirement is that requiring juries to parse unvetted scientific information and disregard those aspects of it they consider suspect is likely to lead to ill-considered verdicts.¹⁰

In contrast to the usual preliminary fact determination, however, the complex nature of scientific evidence has created substantial confusion among courts about just where the judge's authority to decide admissibility ends and the jury's responsibility to assess weight begins. The key variable in the latter setting, according to *Daubert*'s original formulation, is whether the fact is a “conclusion” or something else. Under *Daubert*, in assessing admissibility, the judge's “focus . . . must be solely on principles and

character evidence and related matters, the judge need merely find that “the jury could reasonably find the conditional fact . . . by a preponderance of the evidence.” *Huddleston v. United States*, 485 U.S. 681, 690 (1988).

⁵ See *infra* text accompanying notes 122–23.

⁶ In this Article, we restrict our analysis to scientific evidence. Although we think the points made here are applicable to all expert evidence, we limit the discussion to “scientific” expert testimony because we rely on the structure of scientific inference to set the boundary between admissibility and weight.

⁷ As Jennifer Mnookin has pointed out, however, we do not know much about expert admissibility decisions before the twentieth century. Jennifer L. Mnookin, *Scripting Expertise: The History of Handwriting Identification Evidence and the Judicial Construction of Reliability*, 87 VA. L. REV. 1723, 1827 n.264 (2001) (“[T]here has been little effort to grapple directly with how judges made admissibility determinations about experts in the nineteenth century.”).

⁸ 509 U.S. 579 (1993).

⁹ *Id.* at 589 (“[U]nder the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.”).

¹⁰ See *infra* text accompanying notes 122–26.

methodology, not on the conclusions that they generate.”¹¹ This distinction between methodology and principles on the one hand and case-specific conclusions on the other (which we will refer to as the methodology–conclusions distinction for ease of reference)¹² has assumed major significance. While the Court abandoned the distinction just four years later in *General Electric Co. v. Joiner*,¹³ and while the amendments to Rule 702 three years after *Joiner* made no mention of it,¹⁴ many courts continue to put considerable emphasis on whether the preliminary fact is about “methodology” or instead describes “conclusions.”¹⁵ Only a minority of courts have required that the judge preliminarily determine that the expert’s conclusion was reliably reached using a reliable methodology.¹⁶ Most courts hold that the judge’s sole concern is whether the expert followed an acceptable methodology, and other decisions have even punted some types of methodological issues to the jury.¹⁷

In this Article, we argue that the methodology–conclusions distinction has no principled basis in science and thus should have none in law. Since the distinction does not align with the nature of the evidence that scientists proffer in court, it is destined to fail and should be explicitly jettisoned. In its place, courts should adopt a framework that is consonant with the structure of science itself.

That structure has one central characteristic: science is general in nature, because it involves study of categories of individuals or cases, rather than study of a single individual or case.¹⁸ Generalization permeates the scientific enterprise, cutting across methodology, principles, and conclusions. To take just one example, every aspect of the science on the

¹¹ 509 U.S. at 595.

¹² We also use this shorthand because it is the distinction most often stressed by the courts, *see infra* text accompanying notes 64–66, and because it is the clearest. Whereas the difference between a method and a principle or a method and a conclusion is fairly intuitive, the difference between a principle and a conclusion is not. A principle that is applicable to a particular case can also easily be viewed as a conclusion about a particular case (e.g., the principle that eyewitnesses tend to be poor at cross-racial identifications can also serve as a conclusion for an expert on eyewitness testimony). In any event, as will become clear in this Article, we think that the distinction between principles and conclusions is irrelevant in terms of defining the judge’s role in determining admissibility.

¹³ 522 U.S. 136, 146 (1997).

¹⁴ *See* FED. R. EVID. 702(c) (stating that “testimony,” which presumably includes conclusions, must be based on “reliable principles and methods”).

¹⁵ *See infra* notes 64–65 and accompanying text.

¹⁶ *Infra* notes 64–65 and accompanying text.

¹⁷ *See infra* note 66 and accompanying text.

¹⁸ We describe this aspect of science in David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417, 419 (2014) (“Scientists, in their professional lives, almost invariably measure phenomena at the group level and describe their results statistically.”).

accuracy of eyewitnesses is group-driven: its methodology (for example, comparing experimental and control groups of eyewitnesses exposed to different stimuli), the results derived from the methodology (for example, the finding that people have more difficulty identifying members of another race than members of their own race) and its legally relevant conclusions (for example, an opinion that cross-racial identifications are more suspect than other identifications, all else being equal).¹⁹

As we explained in a previous article, the generality of scientific evidence means that methods and findings that have relevance to one legal case will be relevant in other cases as well.²⁰ Just as legal procedures and principles apply across numerous cases, the methodology and conclusions associated with the research on which experts rely can help resolve numerous disputes. The studies on cross-racial identification that an eyewitness expert describes in a case that takes place in 2015 will have implications for cases decided in 2016 and beyond.

That insight has an important consequence for the distinction between admissibility and weight: scientific procedures and principles, *as well as* any conclusions of general application that are derived from them, ought to be evaluated by judges, not by juries that sit on a single case. It is a well-established aspect of our modern jury system that, while laypeople are in charge of finding facts specific to the case at hand, courts are the appropriate entity for ascertaining legal rules that will have application to other cases,²¹ a practice that is based in part on an assessment of the relative capacities of judges and juries and in part on a desire for uniformity across cases.²² For reasons we develop in this Article,²³ the same rule should apply for *facts* that will have application to other cases, for the same reasons. A determination of whether scientific testimony is reliable should be the province of a legally trained individual cognizant of the difficulties of determining scientific validity and required to make a public ruling subject

¹⁹ For a description of research on eyewitness accuracy, see 2 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW & SCIENCE OF EXPERT TESTIMONY, §§ 15:11–43, Westlaw (database updated Dec. 2015).

²⁰ Faigman, Monahan & Slobogin, *supra* note 18, at 424 (“[S]cience is a product of research that applies generally to all similarly situated cases.”).

²¹ Matthew P. Harrington, *The Law-Finding Function of the American Jury*, 1999 WIS. L. REV. 377, 377 (explaining that, under our current system, judges determine the law and juries apply it, although also noting that in the eighteenth and nineteenth centuries many jurists believed that the jury was to find both fact and law).

²² See Stephen A. Weiner, *The Civil Jury Trial and the Law–Fact Distinction*, 54 CAL. L. REV. 1867, 1875, 1924–25 (1966) (arguing that, unlike jury verdicts, judicial conclusions create “a precedent influencing the determination of future cases presenting a repetition of the historical facts to which the law has been applied”).

²³ See *infra* Section III.A.

to appellate review, not hidden within an isolated verdict delivered by laypeople. Acceptance of that proposition means that the role of the judge and jury should depend not on a distinction between methodology and conclusion but on the distinction between the general and the specific. The reliability that *Daubert*, *Joiner*, and Rule 702 all require the judge to determine as a preliminary fact entails assessing every aspect of scientific evidence, not just its methodology or some other subset of the testimony.

At the same time, the general–specific distinction that derives from the nature of scientific inference also means that, whether they involve methods or conclusions, factual disputes that relate solely to the case at hand are for the jury to assess. Thus, whether an expert in the instant case actually applied the methodology that the judge found valid generally is a matter of weight, as is any conclusion the expert reaches that is applicable only to the litigants. The only caveat here is the traditional one that the judge may keep these issues from the jury when no rational jury could credit the expert’s assertions about them.²⁴

Hence, for example, in the controversial area of “shaken baby syndrome,”²⁵ an expert’s assertion that research indicates that subdural hematoma, retinal bleeding, and brain swelling are indicative of child abuse is a general proposition and a matter of admissibility; an expert’s assertion that the victim in the case had this triad of symptoms, however, is a case-specific assertion and thus a matter of weight.²⁶ Likewise, in a case involving testimony about DNA, the assertion that a particular method of testing DNA is reliable is of general import and a matter of admissibility, while an opinion regarding a case-specific fact, such as whether the technician properly labeled the samples prior to performing the analysis, is a matter of weight. An expert economist in an employment discrimination case who admittedly fails to control for a key variable such as seniority or wage structure in a regression analysis has committed a general error that should lead to exclusion by a judge; but determining whether the economist

²⁴ As we discuss in more detail below, *see infra* text accompanying notes 122–24, evidence codes require that even preliminary facts that go to the jury must still meet a “conditional relevance” threshold, such that a reasonable trier of fact could find that the fact is true. *E.g.*, FED. R. EVID. 104(b).

²⁵ Compare Sandeep K. Narang et al., *A Daubert Analysis of Abusive Head Trauma/Shaken Baby Syndrome—Part II: An Examination of the Differential Diagnosis*, 13 HOUS. J. HEALTH L. & POL’Y 203, 207 (2013) (finding that the science in this area is fundamentally sound), with Keith A. Findley et al., *Shaken Baby Syndrome, Abusive Head Trauma, and Actual Innocence: Getting it Right*, 12 HOUS. J. HEALTH L. & POL’Y 209, 213 (2012) (finding that the science in this area is fundamentally unsound).

²⁶ *See generally* Deborah Tuerkheimer, *The Next Innocence Project: Shaken Baby Syndrome and the Criminal Courts*, 87 WASH. U. L. REV. 1 (2009) (exploring the application of the general science of diagnosing shaken baby syndrome to specific cases involving defendants charged with shaking babies to death). *But see* Joëlle Anne Moreno & Brian Holmgren, *Dissent into Confusion: The Supreme Court, Denialism, and the False “Scientific” Controversy Over Shaken Baby Syndrome*, 2013 UTAH L. REV. 153.

who asserts such a variable was included in the analysis did in fact include it, or whether he or she obtained accurate information about the variable, is an assessment that should be carried out by the jury.

In addition to its consistency with the nature of scientific evidence, this approach to the admissibility–weight issue in expert testimony cases has three important benefits. First, it aligns squarely with the purposes of evidence codes and the Constitution’s due process and right to jury provisions by making optimal use of the relative competencies of judges and juries.²⁷ Juries will be prevented from hearing unreliable evidence, and at the same time (assuming the expert testimony is admitted) they will be given full authority to decide facts relevant to the case before them. Second, the division suggested by the structure of scientific inference implements the key goal—again one echoed, albeit faintly, in both constitutional and evidentiary jurisprudence—of ensuring uniformity between cases regarding general propositions of science.

Finally, this approach has the benefit of clarity. The current focus on methodology–conclusions leaves courts in a quandary because, as a scientific matter, methodology and conclusions are not separable. The alternative we propose is more straightforward: when the statement of fact (or inference) that is asserted to support proffered expert opinion transcends the instant case, it is a preliminary fact to be decided by the court under Rule 104(a). When the statement of fact (or inference) that is asserted to support proffered expert opinion is pertinent only in the instant case (after a judge has found that it is the product of reliable principles and methods), it is a question of weight and only subject to review under Rule 104(b) as a matter of conditional relevance.

More formally, therefore, we propose the following test to identify the boundary between a judge’s obligation to determine admissibility and the jury’s task to assess weight:

Preliminary facts that describe the principles or methods of scientific research or generally applicable conclusions drawn therefrom are within the judge’s responsibility to decide as a matter of admissibility under Federal Rule of Evidence 104(a) and equivalent state rules. Conditional facts that describe whether the expert adhered to a reliable principle or method are matters of weight and within the province of the trier of fact to decide if a reasonable trier of fact could find that the fact is true, as provided in Federal Rule of Evidence 104(b).

This preliminary fact test requires the judge to evaluate whether the expert’s conclusions validly derive from reliable research methods and

²⁷ For a discussion of constitutional considerations, see *infra* Section II.A.

principles, leaving as a conditional fact for the jury the assessment of whether methods and principles the judge has found reliable were in fact applied by the expert as he or she claims (and also, by negative inference, allowing the jury to evaluate the credibility of any expert conclusions that are not “generally applicable” but rather specific only to the parties in the case).

Another important implication of a regime that aligns admissibility decisions about expertise with the structure of scientific inference is that appellate courts must also modify their approach to expert evidence. Because trial judge decisions about scientific reliability are general in nature and are relevant to more than just the case before them, the deference accorded trial courts with respect to case-specific trial findings intrinsic to the case is not appropriate in this setting. We argue that, in contrast to current law directing appellate courts to apply the abuse of discretion standard to trial court decisions applying *Daubert*,²⁸ appellate review of such decisions should be more rigorous, akin to how appellate courts review determinations of law made by trial courts.²⁹

This Article establishes the basis for these proposals by describing the origins of the legal distinction between methodology and conclusions in *Daubert* as well as the present state of confusion surrounding that distinction. It then seeks to bring the evidentiary rules in line with the structure of science. Part I provides a brief historical overview of how the issue arose in *Daubert* and the lower courts’ failure to resolve it. Part II explores the values and principles that underlie the judge–jury division regarding fact-finding, both under the Constitution and the rules of evidence. Part III then describes the primary contribution of the Article by explicating the structure of scientific inference and explaining how the roles of the judge and the jury can be aligned with that structure in light of constitutional and evidentiary principles. Part IV observes how these insights about scientific reasoning enhance the authority of the appellate court to override the trial court’s admissibility decision. The Conclusion summarizes our views.

I. THE GENESIS OF THE METHODOLOGY–CONCLUSIONS DISTINCTION

A fundamental component of the law of evidence is the primary obligation of the judge to decide admissibility, an obligation which,

²⁸ *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 139 (1997) (adopting the abuse of discretion standard of review of trial judge determinations under *Daubert*).

²⁹ Several commentators have reached the same conclusion, *see infra* note 194, but we are the first to put forward this rationale, which provides a uniform theory applying to both the trial and appellate contexts.

counterintuitively, very often requires the judge to be a factfinder.³⁰ The hearsay exclusions and exceptions are particularly clear illustrations of this judicial duty. A “coconspirator’s statement” is only admissible if it was “made by the party’s coconspirator during and in furtherance of the conspiracy.”³¹ An “excited utterance” requires, among other things, that “the declarant was under the stress of excitement” that caused the statement to be made.³² And “dying declarations,” among other requirements, must be made under a belief of “imminent” death.³³ Under the federal rules, these are all factual determinations that must be decided by judges under Rule 104(a).

In the landmark *Daubert* decision about scientific testimony, the Court adhered to this conventional approach of defining the judge’s admissibility responsibilities in terms of a preliminary factual inquiry.³⁴ The Court found that Rule 702 required trial courts to evaluate the reliability and validity of the basis for proffered expert testimony.³⁵ The admissibility assessment under Rule 702 thus hinges, the Court stated, on a preliminary fact under Rule 104(a) involving the soundness of the science being offered in court.³⁶ “Faced with a proffer of expert scientific testimony,” the Court declared, “the trial judge must determine at the outset . . . whether the expert is proposing to testify [about] scientific knowledge.”³⁷

The *Daubert* Court’s ruling that scientific validity constitutes a preliminary fact under Rule 702, while not surprising as a general

³⁰ See generally GLEN WEISSEBERGER & JAMES J. DUANE, FEDERAL RULES OF EVIDENCE: RULES, LEGISLATIVE HISTORY, COMMENTARY AND AUTHORITY § 104.2, at 56 (7th ed. 2011) (noting that preliminary questions to be decided by the judge include “the qualifications of a person to be a witness,” “the existence of a privilege,” and “any other question pertaining to the admissibility of some testimony or other evidence (e.g., whether a hearsay statement is admissible under some exception to the hearsay rules)” (quoting FED. R. EVID. 104(a))).

³¹ FED. R. EVID. 801(d)(2)(E).

³² FED. R. EVID. 803(2).

³³ FED. R. EVID. 804(b)(2).

³⁴ 509 U.S. 579, 592–93 (1993) (“Faced with a proffer of expert scientific testimony, then, the trial judge must determine at the outset, pursuant to Rule 104(a) . . . [that] the reasoning or methodology underlying the testimony is scientifically valid and [that] that reasoning or methodology properly can be applied to the facts in issue.” (footnotes omitted)).

³⁵ The *Daubert* Court explicitly refused to equate the “validity test” of Rule 702 that is imposed on federal courts with the test associated with *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), which requires that judges determine whether the scientific technique or method on which scientific testimony is based is “generally accepted” in the particular field from which it comes. *Daubert*, 509 U.S. at 589. However, some jurisdictions continue to subscribe to the *Frye* rule; thus, it is worth noting that general acceptance is also a “preliminary fact” that must be decided as a component of an admissibility determination, and that our analysis would also apply in a *Frye* jurisdiction. See *infra* text accompanying notes 106–13.

³⁶ *Daubert*, 509 U.S. at 592.

³⁷ *Id.*

evidentiary matter, generated a second issue that is largely unique to scientific evidence: What is the proper focus of the validity assessment to be made by judges? In ordinary evidentiary contexts, the preliminary facts judges must find when applying evidentiary rules are plainly defined and unique to the respective case. Whether a statement was made “in furtherance of the conspiracy” or a declaration was uttered “under the stress of excitement” are straightforward factual inquiries and do not have relevance outside of the case at hand. In contrast, the preliminary fact at issue in *Daubert* was whether the methods and principles of years of scientific research and numerous published studies support expert testimony that Bendectin is a teratogen that causes birth defects when ingested by people like the plaintiff’s mother.³⁸ This is not a straightforward factual inquiry or one that arises only in the case at hand. Thus, the precise scope of the preliminary fact or facts judges must decide in connection with scientific evidence is not as easily discerned as in run-of-the-mill cases involving hearsay and the like.

The *Daubert* Court did not delve into this nuance. Rather it simply stated that, for the judge applying Rule 702, the “overarching subject is the scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission.”³⁹ Using language suggesting that the point was obvious, the Court then added, “The focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.”⁴⁰ The Court offered no further elaboration. In particular, it did not differentiate between conclusions that have general application and conclusions relevant only to the parties in the case, despite the fact that the assertion that Bendectin can cause birth defects is fundamentally different from an assertion that the plaintiff’s birth defects were caused by Bendectin.

Courts and commentators intent on limiting the scope of *Daubert* seized on the Court’s distinction between principles and methodology on the one hand and conclusions on the other.⁴¹ At first take, this latter division might seem appealing. On the practical level, the distinction appears to be an administrable means of dividing the responsibility for evaluating scientific evidence: methods are the procedures scientists use to study a phenomenon and conclusions are the facts about the world those methods

³⁸ See *id.* at 582–85.

³⁹ *Id.* at 594–95.

⁴⁰ *Id.* at 595.

⁴¹ See, e.g., Kenneth J. Chesebro, *Taking Daubert’s “Focus” Seriously: The Methodology/Conclusion Distinction*, 15 CARDOZO L. REV. 1745, 1745–53 (1994) (making this distinction).

reveal. On a conceptual level, the methodology–conclusions demarcation seems attractive because it mirrors the procedure–substance distinction that pervades legal analysis and is generally thought to describe the respective duties of the judge and the jury.⁴²

However, this seemingly innocuous sentence in *Daubert* hides a deep conflict with the scientific enterprise. Admittedly, the methodology–conclusions distinction is well recognized in science. Indeed, the conventional scheme for organizing scientific articles is to divide them into sections according to background (i.e., introduction), methods, results, and discussion (i.e., conclusions).⁴³ But both the results and the conclusions drawn from those results are highly dependent on and interactive with the methods used. Even if the methods are reliable, in the sense that they repeatedly produce the same results, the discussion section of a scientific article often recognizes that the findings may be suspect because certain variables were not taken into account, or might have been different had other methods been used, or are only applicable in narrow sets of circumstances or, while statistically significant, indicate only a small overall effect on the dependent variable.⁴⁴ In assessing the usefulness of empirical data for a specific legal purpose, methods, principles, results, and conclusions are irretrievably linked.

For the same reason, the methodology–conclusions dichotomy is inimical to *Daubert*'s own agenda. For instance, if the judge merely has to decide whether a given methodology is valid in the abstract, an expert witness in a toxic tort case who can show he or she relied on well-

⁴² SAUL M. KASSIN & LAWRENCE S. WRIGHTSMAN, THE AMERICAN JURY ON TRIAL: PSYCHOLOGICAL PERSPECTIVES 131, 141 (1988) (explaining that juries are treated as “passive recipients of information” and the judge as a “master of ceremonies”).

⁴³ This structure can be found in virtually any article published in the scientific journals *Science*, *Neuron*, and *Law & Human Behavior*, for instance.

⁴⁴ As one example of the interrelationship between statistical methods and legally relevant conclusions, consider the controversy over the Implicit Association Test (IAT), a social psychological instrument designed to measure the strength of associations between concepts (e.g., whites, African-Americans) and evaluations (e.g., good, bad) or stereotypes (e.g., athletic, criminal) that are outside a person's conscious awareness. The IAT has been proffered by expert witnesses for the plaintiffs in employment discrimination class action litigation as evidence of “unconscious bias.” *E.g.*, *Pippen v. State*, No. LACL107038, 2012 WL 1388902 (Iowa Dist. Ct. Apr. 17, 2012), *aff'd*, 854 N.W.2d 1 (Iowa 2014). However, critics argue that “the low IAT–criterion correlations . . . counsel strongly against the assumption that scores on the race and ethnicity IATs reflect individual differences in propensity to discriminate.” Frederick L. Oswald et al., *Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies*, 105 J. PERSONALITY & SOC. PSYCHOL. 171, 187 (2013). Proponents of the IAT counter that only approximately “4% of variance in discrimination-relevant criterion measures is predicted by Black–White race IAT measures,” but argue that “[t]his level of correlational predictive validity of IAT measures represents potential for discriminatory impacts with very substantial societal significance.” Anthony G. Greenwald et al., *Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects*, 108 J. PERSONALITY & SOC. PSYCHOL. 553, 560 (2015).

conducted *in vivo* animal studies about the substance alleged to have caused the plaintiff's illness might be allowed to testify to that conclusion, regardless of whether such studies have a reasonable connection with the expert's opinion.⁴⁵ Moreover, leaving to the jury an assessment of every expert "conclusion"—including those that are generally applicable, such as the expert's opinion in *Daubert* that, as the Court put it, "Bendectin can cause birth defects"⁴⁶—effectively relinquishes the judicial gatekeeper role the Court sought to establish. Both of these results run counter to *Daubert*'s central goal of ensuring the reliability of expert testimony through a judicial gatekeeper.

In any event, the Court quickly decided that the line dividing methodology and conclusions was arbitrary and that employing it largely eviscerated the gatekeeping function set out in *Daubert*. Just four years later, the Court returned to the issue in *General Electric Co. v. Joiner*.⁴⁷ In *Joiner*, the plaintiff claimed that his exposure to polychlorinated biphenyls (PCBs) while employed as an electrician had enhanced the onset of his lung cancer.⁴⁸ The trial court granted summary judgment after excluding the plaintiff's experts, on the ground that their testimony linking PCBs and small-cell lung cancer "did not rise above 'subjective belief or unsupported speculation.'"⁴⁹ The United States Court of Appeals for the Eleventh Circuit reversed, finding that the lower court had erroneously "excluded the experts' testimony because it 'drew different conclusions from the research than did each of the experts.'"⁵⁰ According to the Court of Appeals, "a district court should limit its role to determining the 'legal reliability of proffered expert testimony, leaving the jury to decide the correctness of competing expert opinions.'"⁵¹ The Eleventh Circuit thus relied on the methodology–conclusions distinction for assigning admissibility and weight.

On appeal, the principal issue addressed by the Supreme Court was the standard of review governing appellate court analysis of lower court

⁴⁵ See Chesebro, *supra* note 41, at 1749.

⁴⁶ 509 U.S. 579, 583 (1993) (stating that the plaintiff's expert "had *concluded* that Bendectin can cause birth defects" and then describing the studies on which "[t]heir *conclusions*" were based (emphasis added)).

⁴⁷ 522 U.S. 136 (1997).

⁴⁸ *Id.* at 139–40 ("The suit alleged that his exposure to PCB's 'promoted' his cancer; had it not been for his exposure to these substances, his cancer would not have developed for many years, if at all.").

⁴⁹ *Id.* at 140 (quoting *Joiner v. Gen. Elec. Co.*, 864 F. Supp. 1310, 1326 (N.D. Ga. 1994)).

⁵⁰ *Id.* at 141 (quoting *Joiner v. Gen. Elec. Co.*, 78 F.3d 524, 533 (11th Cir. 1996)).

⁵¹ *Id.* at 141 (quoting *Joiner*, 78 F.3d at 533).

admissibility decisions, which the Court set at “abuse of discretion”⁵² (a subject to which we return in Part IV). However, the Court also invested considerable ink discussing the admissibility of the testimony proffered by the plaintiff’s scientific experts. The plaintiff, operating under the assumption that *Daubert*’s methodology–conclusions distinction controlled the extent of the gatekeeping function, had argued in the lower courts that the animal and epidemiological studies upon which his experts relied were reasonable methods for reaching the conclusion that PCBs had enhanced the onset of the plaintiff’s cancer.⁵³ The Supreme Court registered considerable chagrin over this claim, noting that, in effect, it would make admissibility hinge on the validity of the experts’ methods in the abstract rather than as they apply to the case at hand. For example, in regard to the animal studies, the Court complained that the plaintiff “chose ‘to proceed as if the only issue [was] whether animal studies can ever be a proper foundation for an expert’s opinion.’”⁵⁴ The majority categorically rejected this pinched view of the gatekeeping function, stating: “Of course, whether animal studies can ever be a proper foundation for an expert’s opinion was not the issue. The issue was whether *these* experts’ opinions were sufficiently supported by the animal studies on which they purported to rely.”⁵⁵ For the same reason, the Court also agreed with the district court’s determination that the cited epidemiological studies failed to support the experts’ conclusions.⁵⁶

Even more tellingly, the Court backtracked from the language in *Daubert* setting out the methodology–conclusions line of demarcation. According to the majority:

[C]onclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.⁵⁷

While the decision in *Joiner* thus unambiguously dismissed the notion that methodology and conclusions are easily separable, it did not explicitly

⁵² *Id.* at 141, 146.

⁵³ *See id.* at 140.

⁵⁴ *Id.* at 144 (alteration in original) (quoting *Joiner*, 864 F. Supp. at 1324).

⁵⁵ *Id.*

⁵⁶ *See id.* at 145–46.

⁵⁷ *Id.* at 146.

hold that courts are *required* to recognize their nexus.⁵⁸ Nor did it suggest a concrete substitute for the methodology–conclusions demarcation. In 2000, however, the amendments to Rule 702, meant to implement *Daubert* and *Joiner*,⁵⁹ sent a clearer message. The new Rule 702, restyled in 2011, requires not only that expert testimony be helpful, as the old rule did, but also directs the judge to find that “the testimony is based on sufficient facts or data,” “that the testimony is the product of reliable principles and methods,” and that the expert “has reliably applied the principles and methods to the facts of the case.”⁶⁰

This language codifies *Joiner*’s command that courts focus on “whether *these* experts’ opinions were sufficiently supported” by the research.⁶¹ It eschews the methodology–conclusions distinction—as well as any distinction between these two aspects of the testimony and principles—by not only requiring a determination that the expert’s methods and principles are reliable but also that they are reliably applied “to the facts of the case.”⁶² Bringing home this point, the Advisory Committee note to Rule 702 in the 2000 revisions points out that, while the fact that experts disagree about a scientific issue should obviously not be automatic grounds for exclusion, a conclusion by an expert that is at odds with those reached by most others in the field can be a reasonable indication that the expert has not reliably applied the relevant principles and methods.⁶³ The new Rule’s language and the Advisory Committee’s note make clear that the conclusions the expert reaches about the case—at least those that apply generally to other cases—must be found reliable to be admissible.

⁵⁸ The Court did not directly address the issue but rather insisted, throughout the opinion, that the trial judge has significant discretion in making the admissibility decision.

⁵⁹ The rules were also meant to codify the third case in what has been called the *Daubert* trilogy, *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), which applied Rule 702 to all types of expert testimony. See Weissenberg & Duane, *supra* note 30, § 702.4, at 458 (stating that the requirements noted in the text “were added to the Rule in 2000, codifying the Supreme Court’s holdings in *Daubert* . . . and its progeny” (footnotes omitted)).

⁶⁰ FED. R. EVID. 702.

⁶¹ 522 U.S. at 144.

⁶² FED. R. EVID. 702(d).

⁶³ The Advisory Committee first quotes *Joiner*’s statement that “conclusions and methodology are not entirely distinct from one another” and then states that “when an expert purports to apply principles and methods in accordance with professional standards, and yet reaches a conclusion that other experts in the field would not reach, the trial court may fairly suspect that the principles and methods have not been faithfully applied.” FED. R. EVID. 702 advisory committee’s note to 2000 amendment (first quoting *Joiner*, 522 U.S. at 146; and then citing *Lust v. Merrell Dow Pharm., Inc.*, 89 F.3d 594, 598 (9th Cir. 1996)).

Yet not all lower courts have received the message. While some courts have taken to heart the change in focus signaled by *Joiner* and Rule 702,⁶⁴ many other courts, perhaps most, continue to insist on the methodology–conclusions distinction when determining whether an expert evidentiary proposition goes to admissibility or weight.⁶⁵ Still others, especially at the state level or in certain types of cases, appear to be oblivious to the entire issue.⁶⁶

The lower courts are not entirely at fault for this confusion. Given the complex nature of scientific evidence, an unambiguous dividing line that transcends the methodology–conclusions distinction is not immediately apparent. Nor has the Supreme Court, the Federal Rules, or commentators yet offered a clear conceptual basis for making the necessary distinctions. The rest of this Article seeks to remedy this situation.

II. PREMISES THAT INFORM THE ADMISSIBILITY OF SCIENTIFIC EVIDENCE

The beginning of the inquiry into the proper role of judge and jury in scientific evidence cases requires a recognition of two fundamental aspects

⁶⁴ See, e.g., *Savage v. Union Pac. R.R.*, 67 F. Supp. 2d 1021, 1028 (E.D. Ark. 1999) (“*Daubert*’s standard of admissibility ‘extends to each step in an expert’s analysis all the way through the step that connects the work of the expert to the particular case.’ Thus, if the expert’s conclusion—or any inferential link that undergirds it—fails under *Daubert* to provide any evidence of causation, it must be excluded” (citation omitted) (quoting *In re Conrail Toxic Tort FELA Litig.*, Nos. CIV. A 94-11J, CIV. A 94-4J, 1998 WL 465897, at *6 (W.D. Pa. Aug. 4, 1998))).

⁶⁵ See, e.g., *City of Pomona v. SQM N. Am. Corp.*, 750 F.3d 1036, 1047–48 (9th Cir. 2014); *Schultz v. Akzo Nobel Paints, LLC*, 721 F.3d 426, 431 (7th Cir. 2013) (“[T]he key to the gate . . . is the soundness and care with which the expert arrived at her opinion: the inquiry must ‘focus . . . solely on principles and methodology, not on the conclusions they generate.’ So long as the principles and methodology reflect reliable scientific practice, ‘[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.’” (second ellipsis in original) (citation omitted) (quoting *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 595–96 (1993))); *Baker Valley Lumber, Inc. v. Ingersoll-Rand Co.*, 813 A.2d 409, 416 (N.H. 2002) (“The proper focus for the trial court is the reliability of the expert’s methodology or technique. The trial court functions only as a gatekeeper, ensuring a methodology’s reliability before permitting the fact-finder to determine the weight and credibility to be afforded an expert’s testimony.” (citing *Daubert*, 509 U.S. at 592–95)); see also David E. Bernstein & Eric G. Lasker, *Defending Daubert: It’s Time to Amend Federal Rule of Evidence 702*, 57 WM. & MARY L. REV. 1, 7 (2015) (“Although many courts have faithfully applied amended Rule 702, the same divisions that existed in the courts prior to 2000 continue to exist today—and on the very same issues that the Judicial Conference sought to resolve.”).

⁶⁶ See, e.g., *J.H.H. v. State*, 897 So. 2d 419, 422 (Ala. Crim. App. 2004) (“*Daubert* does not require the accuracy of the testing in the particular case to be assessed at the admissibility stage.” (quoting *Turner v. State*, 746 So. 2d 355, 360 (Ala. 1998))); *State v. Pesqueira*, 333 P.3d 797, 802 (Ariz. Ct. App. 2014) (“[Q]uestions about the accuracy and reliability of a witness’ factual basis, data, and methods go to the weight and credibility of the witness’ testimony and are questions of fact” (quoting *Pipher v. Loo*, 212 P.3d 91, 96 (Ariz. Ct. App. 2009))); see also Bernstein & Lasker, *supra* note 65, at 8 (“[A] number of courts have simply ignored the Rule 702 amendment, relying instead on *Daubert* case law prior to the amendment or even on case law prior to *Daubert* itself.”).

of our legal system. First, as both the common law and all evidence codes make clear, judges are the principal arbiters of when evidence is admissible.⁶⁷ Second, most courts and evidence scholars agree that, in carrying out this role and interpreting the multitude of provisions that implement it, judges should be governed by one simple principle: all relevant evidence should be heard by the jury unless there is a good reason to keep it out, such as a concern that it is unfairly prejudicial, misleading, a waste of time, deceptive, redundant, or unreliable.⁶⁸

At the start, then, it is necessary to identify how this basic principle informs the law's reception of scientific evidence in the courtroom. When, if ever, may the judge keep such evidence from the jury? Here we look first at what the Constitution has to say about this issue, and then examine the various approaches found in evidence jurisprudence. Both sources provide some insight into the relative role of judges and juries in scientific evidence cases. But ultimately neither source provides satisfactory guidance on the central concern we are addressing.

A. Constitutional Considerations

The Rules of Evidence must be interpreted against the basic guarantees of the Constitution. The Fifth and Fourteenth Amendments guarantee due process for those whom the government seeks to deprive of life, liberty, and property—language that has been interpreted to require the government to treat litigants, and especially criminal defendants, fairly.⁶⁹ The Sixth and Seventh Amendments guarantee the right to a jury trial in criminal and civil cases, respectively,⁷⁰ and the Sixth Amendment also guarantees a criminal defendant the right “to be confronted with the witnesses against him” and “to have compulsory process for obtaining witnesses in his favor.”⁷¹ This second set of provisions states that criminal

⁶⁷ See, e.g., FED. R. EVID. 104(a) (“The Court must decide any preliminary question about whether . . . evidence is admissible.”).

⁶⁸ See, e.g., FED. R. EVID. 402 (“Relevant evidence is admissible, unless any of the following provides otherwise: the United States Constitution; a federal statute; these rules; or other rules prescribed by the Supreme Court.”).

⁶⁹ See U.S. CONST. amend. V (“No person shall . . . be deprived of life, liberty, or property, without due process of law . . .”); *id.* amend. XIV, § 1 (“No state shall . . . deprive any person of life, liberty, or property, without due process of law . . .”).

⁷⁰ The Sixth Amendment, in relevant part, provides that “[i]n all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed.” *Id.* amend. VI. The Seventh Amendment guarantees that “[i]n Suits at common law, where the value in controversy shall exceed twenty dollars, the right by jury shall be preserved, and no fact tried by a jury, shall be otherwise re-examined in any Court of the United States, than according to the rules of the common law.” *Id.* amend. VII.

⁷¹ *Id.* amend. VI.

defendants and civil litigants have a right to have their cases determined by a jury,⁷² and that criminal defendants have a right to rebut and present evidence relevant to their case.⁷³ Thus, with respect to the judge's role in assessing the reliability of scientific evidence, the Constitution could be said to require that the judge monitor carefully the government's evidence, and at the same time ensure that nongovernmental parties, and in particular criminal defendants, can present their full cases to a panel of laypeople.⁷⁴

We will not attempt a full exploration of the extent to which the Supreme Court has adopted this interpretation. Only enough will be said here to make clear that the Court's constitutional decisions have not succeeded in clearly demarcating the role of the judge from the role of the jury. While some of the Court's cases take the view that the Constitution imposes few constraints on expert testimony, others indicate that reliability and the effect of the testimony on lay jurors are constitutionally relevant considerations. The end result is a very blurry picture of how the Constitution affects the division of responsibility between judges and juries in cases involving scientific evidence.

In *Barefoot v. Estelle*,⁷⁵ the Court was confronted with a type of scientific evidence—regarding predictions of violence—that it conceded was error-ridden.⁷⁶ Yet it held that the guarantees of the Due Process Clause and the Eighth Amendment do not prevent the prosecution from submitting such evidence to the jury, even in a capital case.⁷⁷ On the reliability issue, the Court simply stated: “The suggestion that no psychiatrist's testimony may be presented with respect to a defendant's

⁷² See *Duncan v. Louisiana*, 391 U.S. 145, 149 (1968) (recognizing the right to jury trial in state criminal cases). The Seventh Amendment has not been applied to the states, *Curtis v. Loether*, 415 U.S. 189, 192 n.6 (1974), but most states recognize the jury right in civil cases, David A. Anderson, *First Amendment Limitations on Tort Law*, 69 BROOK. L. REV. 755, 793 (2004) (noting that “almost all” states guarantee a right to jury trial in civil cases).

⁷³ *Washington v. Texas*, 388 U.S. 14, 18–19 (1967) (recognizing right to compulsory process in state criminal cases); *Pointer v. Texas*, 380 U.S. 400, 403 (1965) (recognizing right of confrontation in state criminal cases).

⁷⁴ Cf. *Ohio v. Roberts*, 448 U.S. 56, 66 (1980) (holding that the Confrontation Clause requires the prosecution to demonstrate that hearsay has “adequate ‘indicia of reliability’”), *abrogated by Crawford v. Washington*, 541 U.S. 36 (2004); Janet C. Hoefel, *The Sixth Amendment's Lost Clause: Unearthing Compulsory Process*, 2002 WISC. L. REV. 1275, 1278 (arguing, based on an historical analysis of the Confrontation Clause and the Due Process Clause, that “relevant evidence offered by the accused should be admitted as long as the prosecution can test the reliability of that evidence with the tools of the adversary process”).

⁷⁵ 463 U.S. 880 (1983).

⁷⁶ *Id.* at 901 (noting, and accepting *arguendo*, that “[n]either petitioner nor the [American Psychiatric] Association suggests that psychiatrists are always wrong with respect to future dangerousness, only most of the time”).

⁷⁷ *Id.* at 903.

future dangerousness is somewhat like asking us to disinvent the wheel.”⁷⁸ And the majority was also unconcerned about the possibility the jury would misuse the information. Rather, it stated, “We are unconvinced . . . that the adversary process cannot be trusted to sort out the reliable from the unreliable evidence and opinion about future dangerousness, particularly when the convicted felon has the opportunity to present his own side of the case.”⁷⁹

Barefoot implies that the Constitution does not place significant restrictions on scientifically suspect evidence. In *Rock v. Arkansas*,⁸⁰ the Court at first glance appeared to reinforce that stance by holding that the Constitution sometimes bars attempts by the state to *exclude* scientifically weak testimony, at least when it is presented by a criminal defendant. In *Rock*, the defendant wanted to introduce statements she had made under hypnosis.⁸¹ While the case did not involve expert testimony, it did involve the use of an interview method that, the Court itself recognized, scientists consider vulnerable to both conscious manipulation and unconscious production of erroneous statements.⁸² Nonetheless, the Court concluded that, given defendants’ constitutional rights to testify and present witnesses on their behalf, defendants cannot be absolutely barred from presenting such evidence, unless the state can show that it “is always so untrustworthy and so immune to the traditional means of evaluating credibility that it should disable a defendant from presenting her version of the events for which she is on trial.”⁸³

Three observations caution against giving *Rock*’s liberal treatment of suspect science a broad reading, however. First, *Rock* alluded favorably to the existence of “procedural safeguards” associated with the use of hypnosis, suggesting that some methodological constraints are permissible.⁸⁴ Second, the holding is clearly bottomed on the strong protection the Constitution affords criminal defendants;⁸⁵ reasonable restrictions on scientific evidence presented by the prosecution or civil litigants are unlikely to be struck down by the Court. Finally, two subsequent Supreme Court decisions indicate that *Rock* may be limited to

⁷⁸ *Id.* at 896.

⁷⁹ *Id.* at 901.

⁸⁰ 483 U.S. 44 (1987).

⁸¹ *Id.* at 46–47.

⁸² *Id.* at 60 (recognizing “the unreliability that hypnosis concededly may introduce”).

⁸³ *Id.* at 61.

⁸⁴ *Id.* at 60 (after noting the problems with hypnosis, stating “[t]he inaccuracies the process introduces can be reduced, although perhaps not eliminated, by the use of procedural safeguards”).

⁸⁵ See *Chambers v. Mississippi*, 410 U.S. 284, 302 (1973) (“Few rights are more fundamental than that of an accused to present witnesses in his own defense.”).

rules affecting the right of defendants to testify, meaning that rules that restrict *expert testimony* proffered by criminal defendants can still pass constitutional muster.

In the first of these decisions, *United States v. Scheffer*,⁸⁶ the defendant sought to admit the results of a polygraph test that would have supported his testimony at trial that he did not knowingly use drugs.⁸⁷ The government objected that the relevant evidence provisions established a *per se* rule excluding polygraph evidence.⁸⁸ Consistent with the reasoning in *Rock*, the defendant claimed in response that such a rule violated the Fifth and Sixth Amendment rights “to a meaningful opportunity to present a complete defense,” and “to present polygraph evidence to bolster his credibility.”⁸⁹

The *Scheffer* Court rejected the defendant’s argument, finding that “[a] defendant’s right to present relevant evidence is not unlimited, but rather is subject to reasonable restrictions.”⁹⁰ Among other legitimate interests, the Court highlighted the government’s need to ensure that “only reliable evidence is introduced at trial.”⁹¹ The Court emphasized that “the exclusion of unreliable evidence is a principal objective of many evidentiary rules.”⁹² Applying this basic principle to the case at hand, the Court found that the government had reasonably concluded that polygraph tests, as a general matter, were unsound.⁹³ Moreover, the Court noted that the unreliability of the test itself (separate from the issue of its reliability in the *Scheffer* case) was a valid constitutional basis for exclusion.⁹⁴

Similarly, in *Clark v. Arizona*⁹⁵ the Court upheld a state court rule that prohibits criminal defendants from presenting psychiatric opinion testimony on whether the defendant had the *mens rea* for the charged crime

⁸⁶ 523 U.S. 303 (1998).

⁸⁷ *Id.* at 306.

⁸⁸ *Id.* at 306–07 (citing MIL. R. EVID. 707).

⁸⁹ *Id.* at 307 n.3 (citing *Crane v. Kentucky*, 476 U.S. 683, 690 (1986)).

⁹⁰ *Id.* at 308.

⁹¹ *Id.* at 309.

⁹² *Id.*

⁹³ *Id.* at 312 (“[The government’s approach] is a rational and proportional means of advancing the legitimate interest in barring unreliable evidence.”).

⁹⁴ *Id.* (“[T]here is simply no way to know in a particular case whether a polygraph examiner’s conclusion is accurate, because certain doubts and uncertainties plague even the best polygraph exams.”).

⁹⁵ 548 U.S. 735 (2006).

(while still allowing such testimony on the insanity issue).⁹⁶ Quoting from its own precedent, the Court stated:

While the Constitution . . . prohibits the exclusion of defense evidence under rules that serve no legitimate purpose or that are disproportionate to the ends that they are asserted to promote, well-established rules of evidence permit trial judges to exclude evidence if its probative value is outweighed by certain other factors such as unfair prejudice, confusion of the issues, or potential to mislead the jury.⁹⁷

After recounting reasons why testimony from mental health professionals can be speculative, the Court asserted that “these empirical and conceptual problems add up to a real risk that an expert’s judgment in giving capacity evidence will come with an apparent authority that psychologists and psychiatrists do not claim to have.”⁹⁸

In short, even in cases where a criminal defendant proffers the expert testimony, the Constitution permits states to create a division between judge and jury for purposes of determining admissibility and weight. Perhaps because of a desire to avoid constitutionalizing evidence law,⁹⁹ however, the Court’s cases addressing the admissibility of expert testimony provide little guidance on where the line should be drawn. In particular, these cases devote no attention to the possible benefits of basing the division, as we propose, on the extent to which the testimony is based on general propositions. While, as we note in Part III, constitutional decisions in other domains have signaled some appreciation for the general–specific distinction in connection with factual determinations,¹⁰⁰ the Court’s decisions regarding evidence and preliminary fact questions are silent on this point. Rather, these latter decisions have left resolution of this issue to the rules of evidence.

⁹⁶ *Id.* at 779; *see also id.* at 745 (describing the Arizona rule at issue as one established in *State v. Mott*, 931 P.2d 1046 (Ariz. 1997)).

⁹⁷ *Id.* at 770 (ellipsis in original) (quoting *Holmes v. South Carolina*, 547 U.S. 319, 326 (2006)).

⁹⁸ *Id.* at 778. This language stands in remarkable contrast to the holding in *Barefoot*, which ignored Justice Blackmun’s dissent making the identical point in arguing against the admission of expert prediction testimony presented by the prosecution. *See Barefoot v. Estelle* 463 U.S. 880, 929 (1983) (Blackmun, J., dissenting) (“One can only wonder how juries are to separate valid from invalid expert opinions when the ‘experts’ themselves are so obviously unable to do so.”).

⁹⁹ *See Colorado v. Connelly*, 479 U.S. 157, 167 (1986) (“A statement . . . might be proved to be quite unreliable, but this is a matter to be governed by the evidentiary laws of the forum, *see, e.g.*, Fed. Rule Evid. 601, and not by the Due Process Clause of the Fourteenth Amendment. ‘The aim of the requirement of due process is not to exclude presumptively false evidence, but to prevent fundamental unfairness in the use of evidence, whether true or false.’” (quoting *Lisenba v. California*, 314 U.S. 219, 236 (1941))).

¹⁰⁰ *See infra* text accompanying notes 138–43.

B. Evidentiary Considerations

Evidence codes and the decisions that construe them have provided more guidance on the admissibility–weight issue than the Court’s constitutional cases have, but as our earlier account of judicial confusion over the methodology–conclusions distinction revealed, they still leave much to be desired on this score. The primary focus of evidence jurisprudence in this area, as in many others, has been on the relative ability of the judge and jury to decide preliminary facts.¹⁰¹ Where the preliminary fact involves empirical evidence, the need to ensure consistency across cases has also occasionally been a concern, but clearly a secondary one that has not focused on the general–specific distinction we propose.

As Professor Edward Imwinkelried has explained, even when expert testimony is not involved, evidence law has long been conflicted on the proper roles of judge and jury.¹⁰² During colonial times, American judges followed the English practice of having judges make decisions about preliminary facts.¹⁰³ Beginning in the nineteenth century, however, some courts allowed the jury to reconsider the judge’s admissibility decision, at least when the judge admitted the evidence.¹⁰⁴ This practice became more formalized in the second third of the twentieth century when many courts, apparently influenced by Professor Edmund Morgan’s work, permitted most types of preliminary facts to get to the jury under a theory of “conditional relevance.”¹⁰⁵ As applied by these courts, this theory relegated the role of the judge to deciding, often without hearing any proffer from the opponent of the evidence, whether “the foundational evidence has

¹⁰¹ We are aware that the term “preliminary facts” contains some ambiguity. In general, in the evidentiary context it refers to all facts that underlie or are prefatory to admissibility decisions. Under the Federal Rules of Evidence, some of these facts are labeled “preliminary facts” and are controlled by Rule 104(a). As discussed in the text, such facts must be found by a preponderance of the evidence. Other preliminary, or prefatory, facts are treated as “conditional” under Rule 104(b). A judge’s obligation toward conditional facts is limited to ensuring that a reasonable trier of fact could find them to exist. See Edward J. Imwinkelried, *Judge Versus Jury: Who Should Decide Questions of Preliminary Facts Conditioning the Admissibility of Scientific Evidence?*, 25 WM. & MARY L. REV. 577, 579 (1984) (“In the parlance of preliminary factfinding, the issue of a theory’s validity is a question of conditional relevance, in which the judge’s limited role is to determine whether, as a matter of law, the proponent has presented evidence with sufficient probative value to support a rational jury finding that the fact exists.”).

¹⁰² See *id.*

¹⁰³ *Id.* at 584 (“[In colonial times, v]irtually universal agreement existed that the judge was the final arbiter of preliminary fact questions.”).

¹⁰⁴ See JOHN MACARTHUR MAGUIRE, EVIDENCE: COMMON SENSE AND COMMON LAW 221–22 (1947) (describing one such case).

¹⁰⁵ Edmund Morgan, *Functions of Judge and Jury in the Determination of Preliminary Questions of Fact*, 43 HARV. L. REV. 165, 169 (1929) (arguing that “where the relevancy of A depends upon the existence of B, the existence of B should normally be for the jury”).

sufficient probative value as a matter of law to support a rational jury finding of the preliminary fact's existence."¹⁰⁶

The courts' approach to preliminary facts in cases involving scientific evidence followed the same paths. Prior to *Daubert*, the rule in *Frye v. United States*,¹⁰⁷ which focused on the general acceptability of the basis of the expert's testimony,¹⁰⁸ dominated. Under that rule, the judge determined the preliminary fact of general acceptance.¹⁰⁹ This practice was based largely on concern about the jury's ability to handle expert evidence.¹¹⁰ A second, less frequently mentioned, rationale for *Frye* and the strong judicial role regarding its application was concern about consistency across cases. For instance, in *People v. Kelly*,¹¹¹ the decision that adopted *Frye* in California, the California Supreme Court stated that the general acceptance test would "promote a degree of uniformity of decision."¹¹² However, application of the general acceptance test was haphazard and did not seem to depend on a close examination of the scientific opinion's generality in the way we propose.¹¹³

Moreover, beginning in the 1950s, a number of jurisdictions rejected *Frye* on the ground that it was too restrictive and, beginning in the 1970s, some also rejected it on the additional ground that it was inconsistent with the original version of Rules 401, 403, and 702, stances that seemed to permit admission of almost any relevant expert testimony that appeared to assist the factfinder.¹¹⁴ Some courts in these jurisdictions specifically adopted Professor Morgan's conditional relevance approach in the expert evidence context, which in effect meant that the judge's role was limited to determining whether, as a matter of law, a rational jury could find the preliminary fact that formed the basis of the expert's testimony was

¹⁰⁶ Imwinkelried, *supra* note 101, at 594.

¹⁰⁷ 293 F. 1013 (D.C. Cir. 1923).

¹⁰⁸ *See id.* at 1014.

¹⁰⁹ *See* John William Strong, *Questions Affecting the Admissibility of Scientific Evidence*, 1970 U. ILL. L.F. 1, 10–11.

¹¹⁰ *See id.* at 4.

¹¹¹ 549 P.2d 1240 (Cal. 1976).

¹¹² *Id.* at 1244–45; *see also* 1 PAUL C. GIANNELLI & EDWARD L. IMWINKELRIED, JR., *SCIENTIFIC EVIDENCE* § 1.06[a], at 18–19 (4th ed. 2007); 1 MCCORMICK ON EVIDENCE § 203, at 1152 (Kenneth S. Broun ed., 7th ed. 2013).

¹¹³ *See* Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 COLUM. L. REV. 1197, 1211 (1980) ("It is unresolved whether the *Frye* standard requires general acceptance of the scientific technique or of both the underlying principle and the technique applying it.").

¹¹⁴ For an account of the reaction to *Frye* as of 1980, including the impact of the adoption of the Federal Rules of Evidence in 1975, *see id.* at 1228–31 (stating that "*Frye* may be tottering, but has not yet fallen" and describing debates about whether Federal Rule 702 was consistent with or contrary to *Frye*).

scientifically valid.¹¹⁵ The assumption was that juries were capable of assessing any relevant expert evidence that was not overly misleading.¹¹⁶

Of course, the *Daubert* revolution—consisting of *Daubert*, *Joiner*, and the Supreme Court’s decision in *Kumho Tire Company v. Carmichael*,¹¹⁷ which made clear that *Daubert* applied to all varieties of expert testimony—has significantly changed the landscape. Today, in the federal courts and many state courts, the *Daubert* trilogy reigns.¹¹⁸ Even in states that have not adopted *Daubert*, there is greater emphasis on evaluating the reliability of scientific evidence and having judges play a gatekeeping role on that issue.¹¹⁹

It is possible that this rejuvenation of judicial authority over preliminary facts relating to expert testimony was also motivated in part by a desire for consistency of determinations based on similar scientific evidence.¹²⁰ But if so, that goal was clearly secondary. As the discussion in Part I indicated, the overriding impetus behind *Daubert* and its progeny was the belief that the judge is the appropriate authority to evaluate reliability as a preliminary fact because of the jury’s vulnerabilities. For instance, the fear evidenced by the Court in *Daubert* itself was that a jury is likely to be less able than a judge to evaluate whether the methodologies associated with epidemiological studies are reliable and that, even if a jury

¹¹⁵ Imwinkelried, *supra* note 101, at 599 (citing *State v. Kersting*, 623 P.2d 1095, 1099–1100 (Or. Ct. App. 1981)). There is also an argument that the language of Federal Rule of Evidence 901—providing that in order “[t]o satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is”—suggests that the proponent need merely make a plausible case of authenticity. *Id.* at 599–600.

¹¹⁶ The best known defense of this position came from McCormick:

General scientific acceptance is a proper condition for taking judicial notice of scientific facts, but it is not suitable as a determinant of the admissibility of scientific evidence. Any relevant conclusions supported by a qualified expert witness should be received unless there are distinct reasons for exclusion. These reasons are the familiar ones of prejudicing or misleading the jury or consuming undue amounts of time.

MCCORMICK ON EVIDENCE, *supra* note 112, § 203, at 1153–54 (footnotes omitted); *accord* *McKay v. State*, 235 S.W.2d 173, 175 (Tex. Crim. App. 1950).

¹¹⁷ 526 U.S. 137, 147–49 (1999).

¹¹⁸ See David E. Bernstein & Jeffrey D. Jackson, *The Daubert Trilogy in the States*, 44 JURIMETRICS 351, 355–56 (2004) (indicating that, as of 2004, most states had adopted *Daubert* or interpreted their law consistently with *Daubert*); see also, e.g., FLA. STAT. § 90.702 (2015).

¹¹⁹ See Bernstein & Jackson, *supra* note 118, at 355.

¹²⁰ Compare Jay P. Kesan, *An Autopsy of Scientific Evidence in a Post-Daubert World*, 84 GEO. L.J. 1985, 2035 (1996) (“[A] principled validity standard that falls between the relevancy and *Daubert* standards is only likely to lead to more confusion and less uniformity.”), with *Goeb v. Tharaldson*, 615 N.W.2d 800, 814 (Minn. 2000) (suggesting that, compared to *Frye*, *Daubert* had the effect of undermining uniformity).

decides they are not, it may have great difficulty disregarding conclusions about causation that the expert says derive from them.¹²¹

Indeed, even in those jurisdictions that endorsed the jury-friendly conditional relevance approach, commentators and courts agreed that the judicial role with respect to preliminary facts ought to be strongly influenced by whether the jury has the ability to give the preliminary fact in question the weight it deserves. For instance, both Professor Morgan, the rejuvenator of the conditional relevance approach, and the State of California, which aggressively endorsed it,¹²² recognized that, as Professor Imwinkelried puts it, “a critical factor is whether the jury can realistically disregard the proffered evidence after finding the preliminary fact’s nonexistence.”¹²³ In some cases, such as determining whether a document is authentic, the thought was that juries would have no difficulty with this task: the jury would simply ignore a document it considered fake.¹²⁴ In contrast, using the example with which we started this Article, if jury members were allowed to decide the preliminary fact of whether the death of an out-of-court declarant was imminent, they might have great difficulty ignoring the declarant’s statement during their deliberations, even when they decide the declarant’s death was not imminent. The same type of reasoning applies to preliminary facts associated with scientific evidence—especially given their more complicated nature. Thus, courts in conditional relevance jurisdictions had no trouble concluding that jurors are less likely than a judge to recognize signs of scientific unreliability and that, even if they do recognize such signs, they could easily fail to discount the rest of the expert testimony accordingly.¹²⁵

¹²¹ See 509 U.S. 579, 597 (1993) (“We recognize that, in practice, a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations. That, nevertheless, is the balance that is struck by Rules of Evidence . . .”). Additionally, by investing the power over preliminary fact determinations in judges, *Daubert* and Rule 702 are consistent with the Supreme Court’s modern approach of giving trial courts substantial managerial authority over their dockets. See David L. Faigman, *The Daubert Revolution and the Birth of Modernity: Managing Scientific Evidence in the Age of Science*, 46 U.C. DAVIS L. REV. 893, 921 (2013).

¹²² Imwinkelried, *supra* note 101, at 598 (“The California Evidence Code represents the most extreme implementation of the conditional relevance concept.”).

¹²³ *Id.* at 597.

¹²⁴ This is the reasoning behind Rule 1008, which provides that “in a jury trial, the jury determines . . . any issue [concerning] whether: (a) an asserted writing, recording, or photograph ever existed; (b) another one produced at the trial or hearing is the original; or (c) other evidence of content accurately reflects the content.” FED. R. EVID. 1008.

¹²⁵ See Edward J. Imwinkelried, *Trial Judges—Gatekeepers or Usurpers? Can the Trial Judge Critically Assess the Admissibility of Expert Testimony Without Invading the Jury’s Province to Evaluate the Credibility and Weight of the Testimony?*, 84 MARQ. L. REV. 1, 13–15 (2000) (noting that laying the foundation for scientific evidence can often be a prolonged process that jurors will have difficulty ignoring, especially given the degree of concentration required, and that the probabilistic

In short, whether or not a jurisdiction follows *Daubert*, evidence law's approach to the preliminary fact issue has been strongly associated with concern over jury capacity.¹²⁶ Certainly, that worry is an important consideration. Ultimately, however, an omnibus concern about the jury does not provide enough nuance to determine *which* preliminary facts about scientific evidence—whether they be methods, principles, or conclusions—in *which* types of cases ought to be decided by the judge. Given the complexities of scientific evidence and the differing capacities of jurors, attempting to determine whether, in a particular case, specific facets of scientific testimony are “too complicated” or “simple enough” for jury consumption is probably not possible. In any event, even “simple” evidence that is obviously flawed can be hard to disregard if it is powerful enough, as the routine exclusion of coerced confessions,¹²⁷ eyewitness identifications made under unnecessarily suggestive circumstances,¹²⁸ and polygraph results¹²⁹ suggests.

At bottom, while evidence law's focus on the relative capacities of judge and jury is important, it is of minimal help in deciding how to make the division between admissibility and weight with respect to scientific evidence. Judges are clearly meant to be gatekeepers, and juries are clearly meant to be assessors of witness credibility.¹³⁰ But the scope of the gatekeeping function remains fuzzy. Fortunately, another frame, one that takes jury and judicial competencies into account but is driven by the nature of scientific evidence itself, is available.

nature of scientific research is harder to disregard than evidence the reliability of which is categorical in nature).

¹²⁶ It is worthwhile noting that this jury-incapacity rationale for allowing judges to determine the preliminary facts associated with expert testimony undercuts the methodology–conclusions distinction that *Daubert* recognized and that many courts still follow. Every aspect of scientific evidence—whether it is methodology, principles, or conclusions—can be difficult to understand. To the extent the distribution of power between judge and jury considers the ability of the jury to disregard unreliable evidence, the distinction between method and conclusion is ephemeral.

¹²⁷ See *Arizona v. Fulminante*, 499 U.S. 279, 296 (1991) (holding the admission of a coerced confession cannot be harmless error, in part because “confessions have [a] profound impact on the jury”).

¹²⁸ See *Manson v. Brathwaite*, 432 U.S. 98, 113–14 (1977) (holding that “unnecessarily suggestive” identification procedures should be excluded unless found to be reliable).

¹²⁹ See *United States v. Scheffer*, 523 U.S. 303, 314 (1998) (upholding a rule of per se exclusion of polygraph evidence).

¹³⁰ Renée McDonald Hutchins, *You Can't Handle the Truth! Trial Juries and Credibility*, 44 SETON HALL L. REV. 505, 513–18 (2014) (describing how, by the early twentieth century, courts had accepted the notion that juries are responsible for determining witness credibility).

III. ALIGNING THE RULES OF EVIDENCE WITH SCIENTIFIC INFERENCE

The constitutional decisions relevant to the distinction between admissibility and weight do not answer the question we are seeking to address; they merely pose it again as an inquiry into how to balance the court's duty to ensure evidence is reliable and the litigant's rights to a jury determination and voice. Evidentiary jurisprudence has been more helpful, because it is better attuned to reliability concerns and the capacities of the jury. Yet both the Supreme Court's ruling in *Joiner* and the Federal Rules have rejected the only concrete evidentiary proposal devised to date, which provides that admissibility analysis should focus on methodology and principles while the conclusions thereby derived should be a matter for the jury.¹³¹ That distinction has been replaced by Rule 702, a superior but nonetheless still vague provision requiring that judges gauge the reliability of both the methods forming the basis for the testimony and the way in which the conclusions are applied to the case at hand.

We think that Rule 702 expresses the correct approach. But it is not specific enough. Here we propose a more precise method of determining when preliminary facts about scientific evidence should be decided by the judge, one that derives from the general nature of scientific knowledge and how it applies to specific individual disputes. This proposal optimizes reliability analysis, reserves for judges those admissibility issues that are most akin to their role as guardians of the law, and ensures that juries are involved in deciding all factual issues that are directly relevant to the litigants and within their capacity as a lay evaluator of technical evidence.

A. *The Structure of Scientific Evidence*

The central question addressed in this Article is: When is an aspect of expert scientific testimony a preliminary fact to be decided by the judge? We think that this question cannot be answered without some understanding of what the word "fact" means in scientific cases.

That inquiry begins with an examination of the seminal work of Professor Kenneth Culp Davis. Professor Davis identified two kinds of facts—what he termed "legislative facts" and "adjudicative facts"—that he thought helped define the relative roles of judge and jury.¹³² Legislative facts are facts that have relevance to legal reasoning and the fashioning of

¹³¹ See *supra* text accompanying notes 57–63.

¹³² Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402–03 (1942).

legal rules.¹³³ Adjudicative facts are relevant to the resolution of particular cases.¹³⁴ Davis stated that “[a]djudicative facts usually answer the questions of who did what, where, when, how, why, with what motive or intent.”¹³⁵ In contrast, “[l]egislative facts do not usually concern the immediate parties but are general facts which help the tribunal decide questions of law and policy and discretion.”¹³⁶ As Davis stated, judges typically decide questions of legislative fact, and adjudicative facts are usually within the province of the trier of fact.¹³⁷

Davis’s central insight is that the generality of the factual question can have a major impact on the identity of the factfinder. In some contexts, courts have also long understood this point. Consider, for instance, the Supreme Court’s cases dealing with “constitutional facts.” Many of the Court’s most famous cases involved such facts, including *Brown v. Board of Education*,¹³⁸ which found that black school children are negatively affected by segregation,¹³⁹ and *Roe v. Wade*,¹⁴⁰ which addressed when a fetus becomes “viable.”¹⁴¹ When, as in *Brown* and *Roe*, the Court finds constitutionally relevant legislative facts, it almost certainly assumes that other courts will abide by such findings even though they constitute “facts” rather than “law.” A lower court would be regarded as a maverick if it today concluded that African-American school children are not harmed by segregation or if it ignored the Supreme Court’s definition of viability. Indeed, in a rare explicit statement of this principle, Chief Justice Rehnquist, in the 1986 decision in *Lockhart v. McCree*,¹⁴² explained that appellate courts, and particularly the Supreme Court, should not apply the deferential “clearly erroneous” standard to legislative-fact findings by lower courts, since lower courts might reach contrary conclusions on the basis of the same scientific research.¹⁴³

¹³³ *Id.* at 402; see also FED. R. EVID. 201(a) advisory committee’s note (“Legislative facts . . . are those which have relevance to legal reasoning and the lawmaking process, whether in the formulation of a legal principle or ruling by a judge or court or in the enactment of a legislative body.”).

¹³⁴ Davis, *supra* note 132, at 402.

¹³⁵ KENNETH CULP DAVIS, ADMINISTRATIVE LAW TEXT § 7.03, at 160 (3d ed. 1972).

¹³⁶ *Id.*

¹³⁷ See *id.*; see also Davis, *supra* note 132, at 402 (observing that the evidence rules for finding facts that form the basis for creation of law and policy should differ from the rules for finding facts specific to parties in a particular case).

¹³⁸ 347 U.S. 483 (1954).

¹³⁹ *Id.* at 494.

¹⁴⁰ 410 U.S. 113 (1973).

¹⁴¹ *Id.* at 160, 162–64.

¹⁴² 476 U.S. 162 (1986).

¹⁴³ *Id.* at 168 n.3 (citing *Dunagin v. City of Oxford*, 718 F.2d 738, 748 n.8 (5th Cir. 1983) (en banc) (plurality opinion)).

Thus, at least in constitutional cases, the Court appears to recognize that general empirical propositions should apply in similarly situated cases and that judges are obligated to ensure this is the case. But this insight has usually not been applied in the preliminary fact setting, despite the reality that this setting also requires figuring out whether the judge or the jury should be the factfinder.

On the surface, the reason for this oversight is obvious: nothing about the preliminary fact question in the run-of-the-mill, nonconstitutional case is “general” in the sense contemplated by Davis’s definition of legislative facts. Whether a particular declarant’s death is imminent or a particular person is a coconspirator, for instance, is case-specific or, in Davis’s terms, “adjudicative”; it has nothing to do with legal reasoning or policy (as opposed to the rules making imminent death and participation in a conspiracy relevant to hearsay analysis, which, of course, *is* a policy decision). On first view, the same might be said for scientific evidence. For instance, the ultimate fact in *Daubert* was whether the plaintiff’s birth defects were more likely than not caused by his mother’s ingestion of the defendant’s drug.¹⁴⁴ Given the prevailing taxonomy as the Court would have understood it, all expert testimony relevant to this issue would have been deemed “adjudicative.” Nothing in the case had to do with “questions of law and policy and discretion”—Davis’s definition of legislative facts—as those terms are normally used.

Davis was not far off the mark, however. In a series of articles published prior to *Daubert*, one of this Article’s authors and Professor Laurens Walker built on Davis’s work and proposed a vision of courtroom fact-finding that shows how his insights are, in fact, directly relevant to cases like *Daubert*.¹⁴⁵ Monahan and Walker argued that scientific evidence presents a hybrid between legislative and adjudicative facts, a hybrid they called “framework” facts.¹⁴⁶ As Professors Monahan and Walker explained, “a fundamental characteristic of much scientific research is that its relevance has to be understood at two levels of generality, levels that are

¹⁴⁴ 509 U.S. 579, 582 (1993).

¹⁴⁵ See John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477, 488 (1986); Laurens Walker & John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 CALIF. L. REV. 877, 879 (1988); Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559, 570 (1987).

¹⁴⁶ Monahan and Walker used the term “social framework” because they were interested in describing the use of social science research in court. Their approach, however, is broadly applicable to all science used in court. To capture this idea, in our work we have used the more generic term “empirical framework.” See Faigman, Monahan & Slobogin, *supra* note 18, at 423–24.

analytically separate.”¹⁴⁷ On the one hand, scientific knowledge “is a product of research that applies generally to all similarly situated cases, and, on the other, it is relevant to particular cases that might be instances of the general findings.”¹⁴⁸ Thus, applied science, by its nature, begins generally—it transcends any particular courtroom—but in the courtroom it provides a “framework” that must be applied to specific cases. *Daubert*, for example, first presented the general framework question whether scientific research supports a causal link between Bendectin and birth defects and, second—assuming the first question was answered affirmatively—whether the plaintiff’s birth defects were caused by Bendectin.¹⁴⁹

In a recent article entitled *Group to Individual (G2i) Inference in Scientific Expert Testimony*, we continued to explore the evidentiary significance of this inherent division between the general and the specific in applied science.¹⁵⁰ As we explained in that article, “Fundamental differences exist between how scientists describe phenomena as scientists and how trial courts expect scientists to describe those phenomena.”¹⁵¹ On the one hand, scientists “almost invariably measure phenomena at the group level and describe their results statistically.”¹⁵² On the other hand, trial courts “typically consider cases individually and call upon scientific experts to describe their results categorically.”¹⁵³ The challenge of reasoning from group data in science to individual decisions in law is usefully referred to as G2i. We argued that “the ‘G’ component of the G2i analysis (what we call ‘framework’ evidence) is governed by different admissibility standards than expert testimony aimed at addressing the ‘i’ component of that analysis (which we dub ‘diagnostic’ evidence).”¹⁵⁴

Our article did not consider the perhaps more basic question of whether any aspects of framework or diagnostic evidence should be *immune* from admissibility analysis—that is, whether any facets of science should be considered matters of weight rather than admissibility. Here we propose, consistent with the position of both Professor Davis and Professors Monahan and Walker, that because of their general nature, all framework issues should always fall within the judge’s domain and thus be a matter of admissibility under Rule 104(a) and equivalent state rules. Less

¹⁴⁷ *Id.* at 423.

¹⁴⁸ *Id.* at 424.

¹⁴⁹ *See* 509 U.S. at 582–83.

¹⁵⁰ Faigman, Monahan & Slobogin, *supra* note 18.

¹⁵¹ *Id.* at 419.

¹⁵² *Id.*

¹⁵³ *Id.*

¹⁵⁴ *Id.* at 421.

obviously, while diagnostic facts should usually be reserved for the jury, certain types of diagnostic facts should also be subject to gatekeeping by the judge. The dividing line we propose depends entirely on whether the statement of preliminary fact rests on preexisting research of general applicability or instead is intrinsic to the case at hand.

The following sections elaborate on these proposals. They describe four categories of science in an effort to align the requirements of evidentiary rules with the basic nature of scientific reasoning. We propose using these four categories to establish a bright-line division between the judge's responsibility to decide admissibility and the jury's task to decide weight.

B. Categories of Science

The four categories of science described here broadly reflect different levels of scientific work, from highly theoretical to specifically applied. While not every scientist would necessarily describe their discipline in the way we do, we think these four categories are useful heuristics for thinking about the wide varieties of scientific endeavor in a way that can be related to the law's evidentiary demands. We provide an overview of the categories here, followed by a more detailed look.

Category 1, or *Basic Science*, is the systematic study of foundational phenomena without an end product in mind. It is to be distinguished from applied science, which seeks to develop a method, technology, or application that can be used in daily life. Basic science involves the big-ticket ideas in science—black holes in cosmology, general and special relativity in physics, brain function in psychology or neuroscience, and evolution in biology—that are the stuff of Nobel Prizes and similar plaudits.

Category 2, or *Framework Science* (a label that borrows from our earlier work), is applied science that aims at developing general propositions about the world in a way that will have practical impact. This category includes most of what scientists do day-to-day, usually in the shadow of Category 1 theories or hypotheses. For example, Einstein's discovery of relativity, a Category 1 phenomenon, was the start of an explosion of framework science, ranging from the principles underlying global positioning devices to quantum physics.¹⁵⁵

¹⁵⁵ See Jesse Emspak, *8 Ways You Can See Einstein's Theory of Relativity in Real Life*, LIVESCIENCE (Nov. 26, 2014, 11:55 AM), <http://www.livescience.com/48922-theory-of-relativity-in-real-life.html> [<http://perma.cc/FR24-ZTKP>].

Category 3, or *Diagnostic Science* (another label borrowed from our earlier work), is applied science that develops methods for determining whether particular cases are instances of some general scientific (framework) finding. Scientific research on the toxic effects of benzene might demonstrate that at high enough doses it can cause leukemia. This is a Category 2 issue. But creating a scientific methodology that would permit a valid inference regarding whether particular cases of leukemia are attributable to benzene exposure is a Category 3 endeavor. For example, finding a particular set of gene mutations that cause leukemia and that are specifically associated with benzene exposure would be such a diagnostic method.

Finally, Category 4, or *Application of Diagnostic Science*, involves the application of a diagnostic method or theory in a particular case. If sound diagnostic methods exist (i.e., Category 3 considerations have been met), Category 4 concerns whether the scientist used the validated methodology in the case at hand and whether he or she did so properly. This category is not research per se, but rather ensures, as any good scientist would want, that the product of diagnostic research is used in the manner specified by the research.

Again, the reason to develop these four categories is to implement two basic insights, one from science and the other from law. The principal insight from science is G2i, that is, scientists study phenomena at the group level, and the ultimate legal issue is typically whether a particular case is an instance of some relevant phenomenon. From law, the basic insight is that judges have the constitutional and evidentiary responsibility to manage systemic fact-finding, whereas juries are invested with the authority to decide facts particular to the case. If the challenge with scientific evidence is to identify a principled dividing line between the judge's obligation to decide admissibility and the jury's task to assess weight, courts should seek to identify a "cut-line" that inheres in the nature of scientific evidence itself and conforms to the respective obligations of judge and jury.

These four categories of science help establish such a cut-line. Of the four, three involve exclusively general empirical propositions that transcend any one case. The findings of basic science, the conclusions of framework science, and the existence of a diagnostic methodology that can identify particular instances of a relevant phenomenon are all general scientific propositions that extend beyond any individual case. Only the issue of whether a particular methodology was properly applied to a particular case can be truly case-specific, and even here certain issues can transcend the case and thus fall in the judge's bailiwick.

I. Category 1: Basic Science.—Scientists ordinarily begin with a hypothesis or theory about the existence of some phenomenon, which typically occurs at a fairly abstract and indistinct level. This type of research is usually not conducted with the courtroom in mind and has many uses beyond legal ones. The corpus of research on memory and perception, for example, is highly varied and, at least in its earliest forms, had no clear relevance to any courtroom application.¹⁵⁶ Likewise, the early landmark research on DNA had no pretensions to forensic use.¹⁵⁷

Nonetheless, this research might easily form the basis for expert testimony. Consider, for example, one of the foundational bases for modern eyewitness research, the finding that the brain does not operate like a video camera.¹⁵⁸ If expert testimony were offered on the ways that leading or misleading questions can contaminate an eyewitness's account of some event, the expert is likely to discuss basic brain function and the ways that brains encode and retrieve memories, all based on foundational research. Similarly, DNA profiling can be traced to the discovery of the DNA helix. An expert explaining DNA profiling is likely to begin with the foundational premise of the existence of the DNA molecule.¹⁵⁹

The important point for present purposes is that this Category 1 research is of a general nature. In both of these instances, as well as innumerable others, the basic science on which the scientific opinion rests—whether made explicit or left implicit—is a preliminary fact that transcends the particular case. Thus, it ought to be an admissibility consideration.

Furthermore, that conclusion stands whether the expert is describing the results of the research, the principles undergirding them, or the methods used to discover them. *Daubert's* distinction between methods and conclusions is irrelevant here. What is relevant is whether the factual or research premises on which the expert testimony rests transcend the particular case.

¹⁵⁶ See generally Gordon H. Bower, *A Brief History of Memory Research*, in THE OXFORD HANDBOOK OF MEMORY 3 (Endel Tulving & Fergus I.M. Craik eds., 2000) (describing research on memory in the nineteenth century).

¹⁵⁷ See Sarah L. Bunce, Comment, *United States v. Kincade—Justifying the Seizure of One's Identity*, 6 MINN. J.L. SCI. & TECH. 747, 749–53 (2005) (noting that, while DNA was first discovered in 1869, it was not used in litigation until the late twentieth century).

¹⁵⁸ See, e.g., Donna J. Bridge & Joel L. Voss, *Hippocampal Binding of Novel Information with Dominant Memory Traces Can Support Both Memory Stability and Change*, 34 J. NEUROSCIENCE 2203 (2014); Elizabeth F. Loftus & Jacqueline E. Pickrell, *The Formation of False Memories*, 25 PSYCHIATRIC ANNALS 720 (1995).

¹⁵⁹ See generally JAMES D. WATSON, THE ANNOTATED AND ILLUSTRATED DOUBLE HELIX (Alexander Gann & Jan Witkowski eds., 2012) (describing the discovery of the structure of DNA).

2. *Category 2: Framework Science.*—The majority of scientific research involves the exploration of hypotheses that are suggested by, or thought to be consistent with, Category 1 theories. This work occupies Category 2. It is vast, diverse, and frequently proffered as a basis for expert evidence in court. For example, as noted above, Category 1 theories of brain function theorize that memory does not operate like a video camera that stores accurate representations of our experiences in an unalterable database, but rather deteriorates rapidly over time and is easily influenced by subsequent events. This theory has been tested and its parameters specified in a host of areas potentially highly relevant to legal disputes, including eyewitness accuracy, children as witnesses, lie detection, false confessions, and repressed memories.¹⁶⁰ We call this latter type of testimony “framework” testimony because it provides a frame for legally relevant behavior.

Framework research is inherently general and its validity does not depend on the circumstances of a particular case.¹⁶¹ The issues this category addresses, such as whether Bendectin causes birth defects, sleep deprivation contributes to false confessions, trichloroethylene causes cancer, or young children are highly suggestible, transcend individual disputes. The soundness of this research does not depend on locale; it is as “true,” or as “false,” in San Francisco as it is in Nashville or Charlottesville. And, as with basic science, both the methods *and* the conclusions of this type of research fit this description.

It should also be noted that the validity of Category 2 research does not depend on the existence of a Category 1 foundational phenomenon. To be sure, a compelling umbrella phenomenon can help situate specific research findings. But the annals of science are replete with Category 2 framework-type research that does not fit any existing paradigm. For example, research on predictions of violence has few theoretical pretensions but, at least when based on sound research methods, will be sufficiently valid to admit.¹⁶² Similarly, research might show, to a high degree of confidence, that benzene causes a particular form of leukemia,

¹⁶⁰ See FAIGMAN ET AL., *supra* note 19, ch. 15 (eyewitness identifications); *id.* ch. 16 (children’s memory and testimony); *id.* ch. 19 (repressed memories); 5 FAIGMAN ET AL., *supra* note 19, ch. 38 (polygraph tests); Saul M. Kassin et al., *Police-Induced Confessions: Risk Factors and Recommendations*, 34 LAW & HUM. BEHAV. 3 (2009).

¹⁶¹ Of course, its relevance (or fit) might very well depend on the particulars of the case. Research on cross-race identifications might be valid, but it would obviously be irrelevant, and thus inadmissible, in a case that involved a same-race identification.

¹⁶² See John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 405–27 (2006); Christopher Slobogin, *A Jurisprudence of Dangerousness*, 98 NW. U. L. REV. 1, 50–53 (2003).

but scientists might not be able to identify the specific biological mechanism of this relationship.¹⁶³ From the law's perspective, therefore, well-validated Category 2 framework evidence might be admissible even absent a more general theory.

3. *Category 3: Diagnostic Science.*—In some cases the only type of expert testimony presented will consist of Category 1 and Category 2 research, or Category 2 research alone. For instance, expert eyewitness testimony usually reports only general research findings and does not address whether a particular eyewitness is likely to be accurately reporting what was observed.¹⁶⁴ If it did, however, reliability concerns arise not only in connection with its general research basis but also with respect to the method used to apply that research to the case at hand. The latter issue implicates Category 3 inference, which we call diagnostic science, because it is concerned with whether there are methods available, grounded in sound science, that permit determining whether particular cases are instances of some general phenomenon.

As with Categories 1 and 2, the applied methods of Category 3 expert testimony will describe the results of preexisting research or scientific practice and be applicable across cases. For instance, doctors purport to be able to diagnose an individual using specific types of tests or protocols developed and used in past cases.¹⁶⁵ Similarly, psychiatrists purporting to address the mental state of a criminal defendant rely, or should rely, on specific interview techniques that are generally accepted in the profession.¹⁶⁶ DNA experts often assert they can match the defendant's DNA with DNA found at a crime scene with a high degree of certainty based on well-developed methods, such as polymerase chain reaction

¹⁶³ Jac A. Nickoloff et al., *Mechanisms of Leukemia Translocations*, 15 CURRENT OPINION HEMATOLOGY 338, 340 (2008).

¹⁶⁴ See FAIGMAN ET AL., *supra* note 19, § 15:1 (describing typical testimony by eyewitness experts).

¹⁶⁵ See STEDMAN'S MEDICAL DICTIONARY 531 (28th ed. 2006) (defining "differential diagnosis" as "the determination of which of two or more diseases with similar symptoms is the one from which the patient is suffering, by a systematic comparison and contrasting of the clinical findings"); THOMAS B. NEWMAN & MICHAEL A. KOHN, EVIDENCE-BASED DIAGNOSIS 3 (2009) (describing differential diagnosis).

¹⁶⁶ Two examples are the Rogers Criminal Responsibility Assessment Scales (R-CRAS), which purports to assist in evaluations of mental state at the time of an offense, and the MacArthur Competence Assessment Tool for Criminal Adjudication (MacCAT-CA), which purports to assist in evaluating defendants' competency to proceed. See Norman G. Poythress, *MacArthur Competence Assessment Tool for Criminal Adjudication (MacCAT-CA)*, in 2 ENCYCLOPEDIA OF PSYCHOLOGY AND LAW 464 (Brian L. Cutler ed., 2008); Richard Rogers, *Rogers Criminal Responsibility Assessment Scales (R-CRAS)*, in 2 ENCYCLOPEDIA OF PSYCHOLOGY AND LAW, *supra*, at 703.

(PCR) technology.¹⁶⁷ Thus, as with the previous two categories, determining whether Category 3 testimony is valid should also be decided as a matter of admissibility.¹⁶⁸

While Category 2 science might proceed without a corresponding Category 1 overarching theory, Category 3 research cannot exist without the findings of Category 2 framework science. If research fails to demonstrate to a sufficient degree of confidence that a particular Category 2 finding is valid, there can be no method available to identify instances of that finding. Put another way, if the major premise (i.e., Category 2) of a scientific assertion has not been shown to exist, the minor premise (i.e., Category 3) cannot be sound. For example, if research does not support the Category 2 statement that Bendectin can cause birth defects, no methods can exist to demonstrate that particular cases of birth defects are attributable to Bendectin.

At the same time, many areas of science might be well supported at the Category 2 framework level but have little or no basis in Category 3. In other words, framework research might indicate that a particular finding is true in general, but scientists may not have been able to develop a diagnostic methodology that permits valid statements to be made about individual cases. This appears to be the case, for instance, with respect to the accuracy of individual eyewitnesses;¹⁶⁹ if so, eyewitness experts should not be allowed to offer Category 3 testimony.

Indeed, the key insight of G2i is that typically the ultimate case-specific question in the courtroom is not the subject of study in science. Because science is usually limited to exploring group differences and general phenomena, no *scientific* methodology reasoning from group data to individual cases may exist; to put the point another way, the law often asks empirical questions to which scientists have no answers. Thus, as *Joiner* signaled,¹⁷⁰ courts evaluating the admissibility of expert testimony

¹⁶⁷ See, e.g., *United States v. Davis*, 602 F. Supp. 2d 658, 664–67, 677 (D. Md. 2009) (describing PCR technology and how “statistics are used to evaluate how likely it is that a similar match would occur if the DNA sample were drawn randomly from the population” in holding that “there is no basis under *Daubert* or the Federal Rules of Evidence to exclude evidence of the DNA matches in this case”).

¹⁶⁸ Cf. Rebecca Haw Allensworth, *Law and the Art of Modeling: Are Models Facts?*, 103 GEO. L.J. 825, 852 (2015) (“Models and their conclusions . . . are better evaluated by a judge . . .” (footnote omitted)). In Faigman, Monahan & Slobogin, *supra* note 18, we developed an analytical framework for making the admissibility determination with respect to such “diagnostic” testimony. See *id.* at 476–80 (summarizing “best practice guidelines”).

¹⁶⁹ See Brian L. Cutler & Gary L. Wells, *Expert Testimony Regarding Eyewitness Identification*, in PSYCHOLOGICAL SCIENCE IN THE COURTROOM: CONSENSUS AND CONTROVERSY 100, 113 (Jennifer L. Skeem, Kevin S. Douglas & Scott O. Lilienfeld eds., 2009) (“The state of the science . . . does not permit an assessment of the accuracy of an individual eyewitness.”).

¹⁷⁰ See *supra* text accompanying notes 55–58.

purporting to address the case before them need to be very careful in deciding whether the case is an instance of some relevant phenomenon that science has studied.

An example of such caution comes from a series of Supreme Court cases analyzing the relevance of developmental studies and neuroscience to sentencing decisions in juvenile cases. In *Roper v. Simmons*,¹⁷¹ the Court held that the developmental immaturity of adolescents is relevant under the principles of the Eighth Amendment, and that this immaturity is a key reason adolescents as a group should be exempted from the death penalty.¹⁷² In subsequent cases, the Court relied on the same reasoning in concluding that adolescents may not be sentenced to life without parole (LWOP) for nonhomicide offenses and may not receive a mandatory sentence of LWOP for homicide offenses.¹⁷³ The holdings of these cases all depend on a Category 2 determination that, on average, adolescents are sufficiently less developmentally mature that they should be treated differently than adults.

Yet, as a practical matter, this Category 2 framework judgment must be followed by individual sentencing decisions. One can imagine Category 2 science that helps in these situations as well. For instance, research on juvenile development might be able to make broad pronouncements about the relative maturity of 15- to 17-year-olds as opposed to 12- to 14-year-olds and 18- to 20-year-olds.¹⁷⁴ But judges and litigants usually also want the expert to address the maturity of a specific adolescent, not just an age-related category of adolescents. If so, a Category 3 diagnostic question arises. Specifically, is there a valid scientific methodology or test that would permit courts to assess the developmental maturity of individual adolescents?

¹⁷¹ 543 U.S. 551 (2005).

¹⁷² *Id.* at 569–71 (stating, after surveying the developmental literature, that “[o]nce the diminished culpability of juveniles is recognized, it is evident that the penological justifications for the death penalty apply to them with lesser force than to adults”).

¹⁷³ *Graham v. Florida*, 560 U.S. 48, 91–92 (2010) (holding that sentencing a juvenile to life without parole for a nonhomicide offense violates the Eighth Amendment, in part because “culpability or blameworthiness is diminished, to a substantial degree, by reason of youth and immaturity” (quoting *Roper*, 543 U.S. at 571)); *Miller v. Alabama*, 132 S. Ct. 2455, 2465 (2012) (holding that mandatory sentencing of a juvenile to life without parole for a homicide violates the Eighth Amendment, and explaining that “*Roper* and *Graham* emphasized that the distinctive attributes of youth diminish the penological justifications for imposing the harshest sentences on juvenile offenders, even when they commit terrible crimes”).

¹⁷⁴ See, e.g., Elizabeth Cauffman & Laurence Steinberg, *(Im)maturity of Judgment in Adolescence: Why Adolescents May Be Less Culpable than Adults*, 18 BEHAV. SCI. & L. 741, 752, 756 (2000) (presenting data differentiating eighth graders, tenth graders, twelfth graders, and young adults in terms of their capacity to make prosocial judgments).

In *Roper*, the Court in essence concluded that the absence of such a diagnostic methodology required adopting a categorical rule exempting everyone under eighteen years of age from the death penalty.¹⁷⁵ Justice Kennedy conceded that some juvenile offenders possess adult-level maturity.¹⁷⁶ But he added that psychologists cannot reliably identify the members of this subset, stating, “It is difficult even for expert psychologists to differentiate between the juvenile offender whose crime reflects unfortunate yet transient immaturity, and the rare juvenile offender whose crime reflects irreparable corruption.”¹⁷⁷

Of course, the fact that scientists have yet to develop a valid method to determine whether an individual case is an instance of some general phenomenon does not have to mean that the courts must rely solely on Category 1 and 2 evidence and cannot address the individual case at all. Rather, the lack of a Category 3 applied methodology will usually simply require that the courtroom factfinder decide the issue without expert assistance. For example, in noncapital sentencing decisions involving juveniles, adolescent maturity is still very relevant,¹⁷⁸ and judges lacking Category 3 expert testimony can base their assessment of a particular juvenile’s culpability on lay and observational evidence.

In sum, judges need to determine the reliability not only of Category 1 and Category 2 testimony, but of Category 3 diagnostic testimony as well. In doing so, they should keep in mind that many areas of scientific evidence have no methodologies available to assist juries in deciding whether the case before them is an instance of the general phenomenon of interest. In such cases, only Category 1 and 2 expert testimony should be permitted as a means of educating the jury, which is then left to determine the applicability of the general scientific findings to the instant case.

4. *Category 4: Application of Diagnostic Science.*—While science generally is devoted to the G and not the i, there are some areas of scientific evidence—medical and psychiatric testimony were mentioned above—in which experts claim to have developed principles and methods that enable them to assist the jury in saying something about a single case. The previous sections considered whether those principles and methods are valid as a general matter. If they are, then the sole remaining admissibility-weight issue is whether the expert, in the words of Rule 702(d), “has

¹⁷⁵ See 543 U.S. at 573–75.

¹⁷⁶ *Id.* at 574.

¹⁷⁷ *Id.* at 573.

¹⁷⁸ *Miller*, 132 S. Ct. at 2468 (indicating that “immaturity, impetuosity, and failure to appreciate risks and consequences” should be taken into account in sentencing juveniles).

reliably applied the principles and methods to the facts of the case.¹⁷⁹ This is Category 4.

This category, in contrast to the previous three, is relevant only to the case at hand. It should thus generally fall in the jury's domain, not the judge's. However, it is important to distinguish two ways in which application of a diagnostic method could be compromised. The first is when the expert concededly fails to employ an approved diagnostic methodology but instead uses some variation on (or an alternative to) the approved methodology. The second is when the parties dispute whether the approved methodology was followed. In the first instance, the expert is, in essence, claiming that a variation on the approved methodology is sufficiently reliable, which makes the issue a Category 3 diagnostic question that should be heard by the judge. In the second instance, in contrast, the issue is a case-specific dispute over the expert's conduct, and is a matter for the jury.

For example, assume that the PCR technology for DNA analysis discussed earlier has been found valid as a diagnostic method either in the instant case or in previous cases. In our view, the judge should also determine, as a preliminary matter, that the expert claims to have used the PCR test. If instead the expert admits that he or she did not use the PCR test but rather some other test, the expert should not be allowed to testify unless the judge finds that the test employed is also reliable. If, on the other hand, the expert claims he or she used the PCR test and did so in the manner it is intended to be used, the jury ought to be allowed to determine whether the expert is telling the truth about these claims.

This second issue, in contrast to the first one, is entirely case-specific. Whether the expert used the PCR test properly in the instant case is a fact that will apply to no other case. We therefore would draw a distinction within Category 4 between whether the expert used a validated methodology in the case at hand (an issue of admissibility under Rules 702(d) and 104(a)), and whether the expert who claims he or she used the valid methodology the way it is supposed to be used in fact did so (an issue of weight).

We would also impose one significant restriction on the jury's usual prerogative to make final determinations on the Category 4 issue of whether the expert properly used a validated diagnostic methodology: a conditional relevance limitation. Even under that relatively jury-centric approach to scientific evidence, the judge is required to withhold from the jury preliminary facts proof of which is so weak that no rational jury could

¹⁷⁹ FED. R. EVID. 702(d).

conclude they existed. This scenario can occur fairly often with respect to expert testimony. The cases are rife with situations in which the expert clearly did not follow the generally accepted procedure.¹⁸⁰ If a reasonable trier of fact could not find that the expert properly used the validated methodology,¹⁸¹ then the judge should not allow the jury to hear the testimony based on it.¹⁸²

One might well ask why that same concern should not leave for the judge the assessment of *any* preliminary fact that is associated with scientific evidence, even one that is not obviously erroneous, given that such evidence is usually complicated in nature.¹⁸³ However, our position on Category 4 application testimony better reflects the constitutional and evidentiary preference for jury decisionmaking on issues of credibility.¹⁸⁴

¹⁸⁰ This includes *Kumho Tire Co. v. Carmichael*, where the expert did not follow his own stated methodology. 526 U.S. 137, 155–56 (1999) (noting that expert testified that “where there is reason to suspect an abnormal bead groove he would ideally ‘look at a lot of [similar] tires’ to know the grooving’s significance, and that he had not looked at many tires similar to the one at issue” (alteration in original)); *see also* *Mut. Benefit Ins. Co. v. Kaz, Inc.*, No. 1:12-CV-2108, 2014 WL 671445, at *3 (M.D. Pa. Feb. 20, 2014) (differentiating the instant case from one involving a fire investigation where “the expert admitted that he did not follow any fixed set of guidelines”); *U.S. Filter Corp. v. Ionics, Inc.*, 68 F. Supp. 2d 48, 68 (D. Mass. 1999) (finding no credible evidence in support of the expert’s contention because the “defendant’s witness admitted that the sieving procedure used by [the expert] was flawed”); *Brown v. Lifescan, Inc.*, No. 96 C 6215, 1998 WL 42264, at *1 (N.D. Ill. Jan. 28, 1998) (noting that the expert “admitted that his sample was too small to provide an opinion . . . and that he did not follow accepted protocol for testing blood glucose monitors”).

¹⁸¹ Determining whether a methodological step is critical can be a difficult question. *Cf. In re Paoli R.R. Yard PCB Litig.* 35 F.3d 717, 745 n.14 (3d Cir. 1994) (“[I]f a court finds that an expert has employed a methodology only slightly different from a methodology that the court thinks is clearly reliable, the court should be more likely to accept the altered methodology than if it was evaluating that methodology as an original matter.”).

¹⁸² *See, e.g., Carlson v. Okerstrom*, 675 N.W.2d 89, 105 (Neb. 2004) (“When a step in an otherwise valid methodology is performed incorrectly, we fail to see how the expert’s results can be any more reliable than if the methodology itself had been wholly invalid. Accordingly, we hold that it is not enough for the trial court to determine that an expert’s methodology is valid in the abstract. The trial court must also determine if the witness has applied the methodology in a reliable manner.”). One method of performing this judicial filtering role is to require the expert to provide an affidavit that the methodology was carried out properly. Another is to require that the affidavit include “facts that both agree and disagree with [the expert’s] opinion—information that will ‘help others to judge the value of [the expert’s] contribution.’” Shelley Storer, *The Weight Versus Admissibility Dilemma: Daubert’s Applicability to a Method or Procedure in a Particular Case*, 1998 U. ILL. L. REV. 231, 251 (second alteration in original) (quoting RICHARD P. FEYNMAN, “SURELY YOU’RE JOKING, MR. FEYNMAN!” 312 (1985)).

¹⁸³ *See, e.g., Imwinkelried, supra* note 125, at 34 (arguing that when there is a “genuine credibility dispute” the judge ought to hear opposing evidence that has “relatively direct relevance to the credibility dispute”).

¹⁸⁴ *See supra* text accompanying notes 142–43. It is also the view of numerous courts. *See, e.g., United States v. Gipson*, 383 F.3d 689, 697 (8th Cir. 2004) (“[W]hen the *application* of a scientific methodology is challenged as unreliable under *Daubert* and the methodology itself is otherwise sufficiently reliable, outright exclusion of the evidence in question is warranted only if the methodology ‘was so altered [by a deficient application] as to skew the methodology itself.’” (alteration in original) (quoting *United States v. Beasley*, 102 F.3d 1440, 1448 (8th Cir. 1996))); *State v. Langill*, 945 A.2d 1, 11 (N.H. 2008) (“Where errors do not rise to the level of ‘negat[ing] the basis for the reliability of the

When the probative value of proffered expert testimony depends on the validity of the methods or statistical analyses employed in the underlying research (Category 3), clues drawn from demeanor or general context are unlikely to provide much insight into a witness's veracity. However, when the factual dispute is case specific—Did the forensic lab staff mix DNA samples?¹⁸⁵ Did the psychologist follow the MMPI protocol? Does the plaintiff have brain damage? Is the defendant suffering from schizophrenia?—demeanor, context, and other intuitive markers are likely to be available to permit the jury to weigh the evidence.

To be sure, research indicates that demeanor is not a very good basis for assessing credibility and that laypersons have difficulty distinguishing truthful statements from deceitful ones.¹⁸⁶ Although these facts might reduce confidence in jury fact-finding, there is no reason to believe that judges are any better.¹⁸⁷ That people may be bad lie detectors is an inherent limitation of our system of trial process. Given the importance of the jury system to that process, the default must be that jurors have the latitude to make the sort of assessments that are presented by case-specific disputes. Jurors have as much, and possibly more, common sense capacity as judges to resolve disputes over whether the technician or other expert in the case did what she said she did.¹⁸⁸ Hence, although judges must ensure that a validated methodology exists and that it was ostensibly employed in the case at hand, the jury should decide whether it was applied reliably, unless the judge finds such clear evidence to the contrary that no rational juror could consider the testimony based on it credible.

principle itself,' the adversary process is available to highlight the errors and permit the fact-finder to assess the weight and credibility of the expert's conclusions." (alteration in original) (citation omitted) (quoting *United States v. Martinez*, 3 F.3d 1191, 1198 (8th Cir. 1993)); *Wise v. Ludlow*, 346 P.3d 1, 15 (Wyo. 2015) ("Differential diagnosis is a reliable methodology. If [the expert] did not correctly follow the methodology of differential diagnosis, that could affect the weight and persuasiveness of her opinions, but does not render that evidence inadmissible under *Daubert*.").

¹⁸⁵ For instance, in studying DNA, four classes of performance errors have been identified: "quirks of nature," honest mistakes, negligence, and fraud, all of which could affect validity. Bert Black et al., *Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge*, 72 TEX. L. REV. 715, 775 (1994). In our view, disputes of this nature should be determined by the jury, based on evidence presented by the parties, unless no rational jury could find an absence of error.

¹⁸⁶ See generally Aldert Vrij, DETECTING LIES AND DECEIT: PITFALLS AND OPPORTUNITIES (2008); Max Minzner, *Detecting Lies Using Demeanor, Bias, and Context*, 29 CARDOZO L. REV. 2557 (2008).

¹⁸⁷ See generally Stephen Porter & Leanne ten Brinke, *Dangerous Decisions: A Theoretical Framework for Understanding How Judges Assess Credibility in the Courtroom*, 14 LEGAL & CRIMINOLOGICAL PSYCHOL. 119 (2009); Sheng Kung Michael Yi et al., *The Wisdom of the Crowd in Combinatorial Problems*, 36 COGNITIVE SCI. 452 (2012).

¹⁸⁸ See generally Charles F. Bond Jr. & Bella M. DePaulo, *Individual Differences in Judging Deception: Accuracy and Bias*, 134 PSYCHOL. BULL. 477 (2008); Barbara A. Spellman & Elizabeth R. Tenney, *Credible Testimony in and out of Court*, 17 PSYCHONOMIC BULL. & REV. 168 (2010).

C. A Test Based On Scientific Inference

Rule 702 requires that testimony be “the product of reliable principles and methods” and also requires that the witness “reliably appl[y] the principles and methods to the facts of the case.”¹⁸⁹ Translating our categories into the Rules’ “principles and methods” language, the results derived from the inferences described in Categories 1 and 2 are best classified under the principles rubric. Both basic research and framework research produce findings setting out general scientific principles. Category 3, in contrast, has more to do with methods, in particular whether the diagnostic methods used by the expert are a reliable means of addressing the specific facts of the case at hand. Finally, Category 4 has to do with both principles and methods, specifically, whether the expert followed a scientifically tested procedure and reached a scientifically helpful conclusion about an issue related solely to the case at hand.

With our categories of science thus translated, we reproduce our proposal for determining when a fact associated with scientific evidence should be decided by the judge as a matter of admissibility or by the jury as a matter of weight:

Preliminary facts that describe the principles or methods of scientific research or generally applicable conclusions drawn therefrom are within the judge’s responsibility to decide as a matter of admissibility under Federal Rule of Evidence 104(a) and equivalent state rules. Conditional facts that describe whether the expert adhered to a reliable principle or method are matters of weight and within the province of the trier of fact to decide if a reasonable trier of fact could find that the fact is true, as provided in Federal Rule of Evidence 104(b).

Again, our argument, based on *Joiner*, is that generalizable scientific conclusions, as well as scientific principles and methods, must be evaluated by the judge. Thus, under this proposal, testimony about scientific research would be admissible only if the judge finds (by a preponderance of the evidence)¹⁹⁰ that both the methods and principles underlying the research *and* all generally applicable conclusions derived from that research are reliable. If that threshold is crossed (and other evidentiary prerequisites are met),¹⁹¹ juries would hear the expert’s testimony unless no rational juror could give it credence. Of course, jurors are free to reassess the validity of

¹⁸⁹ FED. R. EVID. 702(d).

¹⁹⁰ See *supra* note 4.

¹⁹¹ Our proposed rule says nothing about other criteria of admissibility, including fit, helpfulness, and an assessment of whether the prejudicial impact of the evidence outweighs its probative value. The judge must evaluate these issues as well. For proposals as to how the judge should carry out that task, based on the same general-specific distinction made in the text, see Faigman, Monahan & Slobogin, *supra* note 18, at 472–80.

admitted evidence and must ultimately decide how much weight admitted scientific evidence merits. Case-specific conditional facts that underlie expert testimony are necessarily part of that weight determination and are principally within the province of the jury to decide.¹⁹²

IV. IMPLICATIONS FOR APPELLATE REVIEW STANDARDS

The G2i concept has implications not only for the trial court's assessment of scientific evidence in Categories 1 through 4 but also the appellate court's assessment of the trial court's decisions on those matters. Because the first three categories are general in nature, applicable across all cases involving similar scientific issues, the appellate court ought to analyze the trial court's decisions about them in the same way it considers a trial court's decisions about generally applicable legal principles. Earlier we noted how the Supreme Court has signaled that constitutional facts ought to be the province of appellate courts and treated like precedent.¹⁹³ So too here, appellate courts should see their role in scientific evidence cases as monitors of the lower courts' analysis of general scientific principles and guardians of scientific consistency across cases within their jurisdiction.¹⁹⁴

¹⁹² It is possible that the judge might find that, even if the expert properly carried out the procedure, the resulting diagnostic opinion is incredible because it makes too great a leap from the information known to the expert. This determination bears significant similarity to the judgment the court must make under Rule 704(b), which provides that an expert's opinion that reaches the ultimate issue of "whether the defendant did or did not have a mental state or condition that constitutes an element of the crime charged or of a defense" is "for the trier of fact alone," on the ground that testimony that a person is sane or insane is not based on specialized knowledge. FED. R. EVID. 704(b). Although Rule 704 otherwise permits ultimate issue testimony, the Advisory Committee Note states that, even in cases that do not involve psychiatric testimony, the trial court must determine that the subject matter of the testimony presented is "helpful to the trier of fact" and could also be excluded under Rule 403's balancing of probative value against dangers such as undue prejudice. FED. R. EVID. 704 advisory committee's note.

¹⁹³ See *supra* text accompanying notes 138–43.

¹⁹⁴ Others have reached the same conclusion, albeit without referencing how it aligns with the nature of scientific research. See, e.g., *In re Commitment of Simons*, 821 N.E.2d 1184, 1189 (Ill. 2004) ("The trial court's *Frye* analysis . . . is now subject to *de novo* review. In conducting such *de novo* review, the reviewing court may consider not only the trial court record but also, where appropriate, sources outside the record, including legal and scientific articles, as well as court opinions from other jurisdictions."); *State v. O'Key*, 899 P.2d 663, 688 n.45 (Or. 1995) (en banc) ("When the preliminary facts are not case-specific, little or no deference to the trial court's finding is appropriate."); Lisa Heinzerling, *Doubting Daubert*, 14 J.L. & POL'Y 65, 81 (2006) ("One of the potential embarrassments of *Joiner's* abuse of discretion standard is the possibility of apparently inconsistent evidentiary judgments among courts. Since one consequence of this lenient standard of review is that district judges may come to different conclusions on the same evidence, it may be that different judges could find [that evidence regarding a theory of causation] is both reliable and unreliable."); Christopher B. Mueller, *Daubert Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers*, 33 SETON HALL L. REV. 987, 988–89 (2003) ("[T]he *Daubert* revolution would achieve more if appellate courts abandoned the abuse-of-discretion standard in reviewing the rulings of trial judges in this area.").

This insight, if accepted, would mean that the decision in *Joiner* would have to be revisited. Although we earlier signaled our agreement with *Joiner*'s requirement that the reliability of both the expert's methods and the expert's conclusions be examined by trial judges, we do not agree with the part of the opinion holding that appellate courts should review trial court opinions about expert opinion admissibility under an abuse of discretion standard.¹⁹⁵ In the course of so holding, *Joiner* specifically rejected the lower court's application of a "stringent" standard of review on the admissibility decision, stating that this standard "failed to give the trial court the deference that is the hallmark of abuse-of-discretion review."¹⁹⁶ In our view, such deference should only be accorded trial court rulings about case-specific facts, not those adopting generalized propositions. As we developed in Part II, a generalized scientific proposition is closer to law than it is to fact and should be treated accordingly. Thus, while decisions about Category 4 should be subject to an abuse of discretion standard, trial court decisions about the other three categories ought to be reviewed more "stringently." If our analogy between legal principles and scientific principles and methods stands, that review should be de novo.¹⁹⁷

Another advantage of this approach to appellate review is that it counters, at least to some extent, the criticism that the *Daubert* revolution gives too much power to trial court judges.¹⁹⁸ This is not just the complaint that juries should be allowed to provide a different viewpoint than judges (which we think is germane only in Category 4 cases), but the observation that judges are not always driven solely by the goal of assuring the evidence rules are followed. For instance, after cataloguing the various tangential and sometimes illegitimate agendas that might influence trial judges making evidentiary decisions in cases involving scientific evidence, one commentator lamented that "judges can be fairly certain that their *Daubert* rulings will not be overturned" and that "[s]uch deference gives trial judges more opportunity to insert their policy opinions into their

¹⁹⁵ See *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 139 (1997).

¹⁹⁶ *Id.* at 143.

¹⁹⁷ This is the appellate review standard that applies to trial court findings of law. See, e.g., *United States v. Sandsness*, 988 F.2d 970, 971 (9th Cir. 1993) ("Matters of law are reviewed de novo."). As the Supreme Court has observed, "*de novo* review tends to unify precedent." *Ornelas v. United States*, 517 U.S. 690, 697 (1996).

¹⁹⁸ See, e.g., Michael H. Gottesman, *From Barefoot to Daubert to Joiner: Triple Play or Double Error?*, 40 ARIZ. L. REV. 753, 775 (1998) (stating that the *Joiner* decision "places too much discretion in the hands of district judges and makes the outcomes of toxic tort cases in federal courts turn on the prejudices of the particular judge rather than on principles of law"); Jeffrey Robert White, *Experts and Judges*, TRIAL, Sept. 1998, at 91, 92 (arguing that *Joiner* will likely lead to an erosion of the jury's fact-finding role).

decisions.”¹⁹⁹ A more stringent appellate review standard—the *de novo* standard we advocate—would curtail that possibility.²⁰⁰

CONCLUSION

The manner in which scientists make inferences has important implications for evidence law. One such implication is that, in all cases involving scientific testimony, the division of responsibilities between judge and jury and between trial and appellate courts should be rethought. Because most science derives from study of groups and applies across more than one case, it has many of the attributes of law and should be treated in the same fashion. Thus, trial judges should assess the reliability not only of the expert’s methods but also of any conclusions reached by the expert, unless they relate solely to the case at hand. At the same time, whether the expert properly applied a reliable method and whether an expert’s conclusion that relates solely to the case at hand should be given credence are matters of weight for a jury—unless the judge finds that no rational juror could credit the expert’s assertions on these issues. Appellate courts should exercise deference toward the trial court’s admissibility decision with respect to the jury’s case-specific determinations, but should apply a stringent review standard to whether the method the expert purported to use was reliable and whether any generally applicable conclusions purporting to be derived from that methodology were reliably reached.

This alignment of the admissibility–weight determination with the nature of scientific inference also comports with constitutional and evidentiary desideratum. It requires that scientific reliability be assessed by the entities—trial and appellate courts—best equipped to do so, while maintaining the role of the jury as the ultimate arbiter of those case-specific facts that require a credibility assessment and could rationally be decided either way. It is simpler to administer than other admissibility–weight frameworks because the four categories of scientific inference it

¹⁹⁹ Krista M. Pikus, Note, *We the People: Juries, Not Judges, Should Be the Gatekeepers of Expert Evidence*, 90 NOTRE DAME L. REV. 453, 472 (2014); see also Allan Kanner & M. Ryan Casey, *Daubert and the Disappearing Jury*, 69 U. PITT. L. REV. 281, 297–98 (2007) (arguing that because of “increasing caseloads, insufficient trial experience, the duty to ‘manage’ cases, and a bias toward industry,” judges are presented with overwhelming incentives to exclude experts and dismiss cases under *Daubert*).

²⁰⁰ While we think the trial court’s reliability decision ought to be subject to stringent review, its decision about other admissibility issues—fit, helpfulness, and prejudicial impact—might be very case-specific and thus entitled to more deference. For instance, some trial courts hold that expert testimony about eyewitnesses only fits when eyewitnesses are the sole “important” evidence proffered by the prosecution. See FAIGMAN ET AL., *supra* note 19, § 15:9. Some courts find testimony about rape trauma syndrome helpful only when the alleged rape victim has acted in a “counterintuitive” manner after the rape. *Id.* § 14:2. While the expert testimony in either of these two scenarios might be “reliable,” the trial court may exclude it for other, highly case-specific reasons.

contemplates can easily be identified by whether they rely on preexisting, general research and practice or instead involve case-specific assessments. Finally, the proposal provides the best court-monitored mechanism for ensuring that courtroom use of science is both sophisticated and consistent across cases.