

2014

Group to Individual (G2i) Inference in Scientific Expert Testimony

David L. Faigman

UC Hastings College of the Law, faigmand@uchastings.edu

John Monahan

Christopher Slobogin

Follow this and additional works at: http://repository.uchastings.edu/faculty_scholarship

Recommended Citation

David L. Faigman, John Monahan, and Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 *University of Chicago Law Review* 417 (2014).

Available at: http://repository.uchastings.edu/faculty_scholarship/1036

This Article is brought to you for free and open access by UC Hastings Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of UC Hastings Scholarship Repository. For more information, please contact marcusc@uchastings.edu.

The University of Chicago Law Review

Volume 81

Spring 2014

Number 2

© 2014 by The University of Chicago

ARTICLES

Group to Individual (G2i) Inference in Scientific Expert Testimony

David L. Faigman,[†] John Monahan^{††} & Christopher Slobogin[‡]

A fundamental divide exists between what scientists do as scientists and what courts often ask them to do as expert witnesses. Whereas scientists almost invariably inquire into phenomena at the group level, trial courts typically need to resolve cases at the individual level. In short, scientists generalize while courts particularize. A basic challenge for trial courts that rely on scientific experts, therefore, concerns determining whether and how scientific knowledge derived from studying

[†] 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; Co-Director, UCSF/UC Hastings Consortium on Law, Science and Health Policy.

^{††} John S. Shannon Distinguished Professor of Law, Professor of Psychology, and Professor of Psychiatric Medicine and Neurobehavioral Sciences, University of Virginia.

[‡] Milton Underwood Professor of Law, Affiliate Professor of Psychiatry, Vanderbilt University.

We thank the members of the MacArthur Foundation's G2i Committee, who provided valuable comments throughout the development of this Article, including Paul Appelbaum, Josh Buckholtz, Andre Davis, Philip Dawid, Nancy Gertner, Peter Imrey, Russell Poldrack, Marcus Raichle, Valerie Reyna, and Tracey Woodruff. We are also indebted to Kenneth Abraham, Joe Cecil, Ed Cheng, Simon Cole, Gary Edmond, Rebecca Haw, Jerry Kang, Jennifer Lauren, Richard Leo, Gregory Mitchell, Jennifer Mnookin, Erin Murphy, Daniel Murrie, Roger Park, Andrea Roth, Michael Saks, Nicholas Scurich, Dan Simon, Bill Thompson, and Laurens Walker for their valuable 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).

groups can be helpful in the individual cases before them (what this Article refers to as “G2i”). To aid in dealing with this challenge, this Article proposes a distinction between two types of expert evidence: framework evidence that describes general scientific propositions and diagnostic evidence that applies the general propositions to individual cases. It then examines the evidentiary implications of that distinction. Most importantly, admissibility standards for expert testimony should differ depending on whether experts are proffering framework or diagnostic evidence. Judicial analysis of “fit,” expert qualifications, testability, error rates, peer review, general acceptance, helpfulness, and other traditional admissibility criteria for expert evidence will often vary, sometimes significantly, based on this distinction. The Article provides general guidelines about the best practices judges should follow in sorting through these considerations. These guidelines will permit courts to manage G2i inferences in a more informed and coherent way than they do currently.

INTRODUCTION	419
I. THE G2I CHALLENGE	421
A. The Legal Structure of G2i—Framework Evidence and Diagnostic Evidence.....	422
B. Managing G2i.....	425
II. THE COURTS AND G2I	426
A. Basic Legal Standards: <i>Frye</i> and <i>Daubert</i>	427
1. <i>Frye</i>	427
2. <i>Daubert</i>	428
B. How Courts Handle the G2i Challenge.....	432
1. Eyewitness testimony.....	432
2. Medical and psychological conditions.....	434
3. Employment discrimination.....	436
4. Forensic identification.....	438
5. Summary.....	439
III. ADMISSIBILITY CRITERIA FOR FRAMEWORK AND DIAGNOSTIC EVIDENCE	440
A. Relevance (“Fit”)	440
1. Framework testimony.....	441
2. Diagnostic testimony.....	443
B. Qualifications	444
1. Framework testimony.....	444
2. Diagnostic testimony.....	446
C. Scientific Validity.....	447
1. Testability.....	448
2. Error rates and adequate standards.....	453
3. Peer review and publication.....	457
4. General acceptance of the method.....	460
5. Same intellectual rigor.....	463
D. Added Value (“Helpfulness”).....	466
1. Framework testimony.....	467

2. Diagnostic testimony.....	468
E. Exclusion on Prejudice Grounds.....	469
1. Framework testimony.....	469
2. Diagnostic testimony.....	470
CONCLUSION: INTEGRATING BEST PRACTICES INTO JUDICIAL ADMISSIBILITY	
JUDGMENTS.....	472
A. Distinguishing Framework and Diagnostic Evidence.....	474
B. The Best-Practice Guidelines Applied.....	476

INTRODUCTION

Fundamental differences exist between how scientists describe phenomena as scientists and how trial courts expect scientists to describe those phenomena. Scientists, in their professional lives, almost invariably measure phenomena at the group level and describe their results statistically. Trial courts, in contrast, typically consider cases individually and call upon scientific experts to describe their results categorically. There thus exists a basic cultural and language gap between what scientists attempt to do as scientists (that is, generalize) and what they are often called upon to do as expert witnesses (that is, particularize). In practice, this gap is bridged by courts every day, since science-based experts testify regularly.¹ But courts are inconsistent in how they bridge this gap. Sometimes experts are limited to testifying to general group-level phenomena, leaving application and conclusions regarding specific cases to jurors.² Other times experts are permitted to apply their knowledge to an individual case and offer opinions on whether it is an instance of the more general phenomenon.³ The problem is that courts rarely provide any explanation for their decision regarding whether experts will be limited to testifying about general research findings or will be permitted to comment on particular cases.

¹ As Justice Stephen Breyer stated in his introduction to the Federal Judicial Center's *Reference Manual on Scientific Evidence* (National Academies 3d ed 2011), "Scientific issues permeate the law." *Id.* at 3.

² See, for example, text accompanying notes 58–64. Throughout this Article we refer to courtroom triers of fact as jurors. Expert-admissibility rules, which are our primary concern, operate somewhat differently in bench trials. See, for example, *Oklahoma v Tyson Foods*, 565 F3d 769, 780 (10th Cir 2009) ("[A] judge conducting a bench trial maintains greater leeway in admitting questionable evidence, weighing its persuasive value upon presentation.").

³ See text accompanying notes 65–72.

This gap between conventional scientific practice and ordinary trial practice involves the challenge of reasoning from group data to decisions about individuals (an analytical process that we designate as “G2i”). The G2i challenge has been largely ignored by virtually all of the participants in the trial process, including courts, lawyers, experts, and scholars. Yet all applied science presents G2i issues. Indeed, all expert evidence, whether based on controlled experimental research or years of experience, presents G2i issues.⁴ Experts testify to such matters as the conditions likely to lead to false confessions, the indicia of schizophrenia, factors that contribute to eyewitness misidentification, the cancer-causing properties of benzene, and thousands more. These are all general—population-based—statements about the empirical world. They are the “G” of G2i and represent the ordinary perspective of most research and most expertise. However, in the courtroom, the operative questions pertain to the particular case at hand, the “i” of G2i: Did the suspect falsely confess? Does the defendant have schizophrenia? Was the witness’s eyewitness identification accurate? Did benzene cause the plaintiff’s leukemia?

In terms of scientific inference, reasoning from the group to an individual case presents considerable challenges and, simply put, is rarely a focus of the basic scientific enterprise. In the courtroom, it is the enterprise. It turns out that the problem of determining the allowable specificity of an expert’s opinion is highly complex and depends on a myriad of legal and scientific considerations. Resolving this problem requires careful balancing of these considerations as they are presented in specific legal contexts. As a result, no simple prescription is available. However, certain guideposts can be identified, ones that recur throughout expert testimony, thus permitting the development of a set of best practices regarding when courts should—and when they should not—allow experts to apply general research data to particular cases.

In this Article, we set out to develop such a set of best practices by which courts might manage the G2i issue. Specifically, our goal is to set forth criteria that will help courts determine

⁴ See *Kumho Tire Co v Carmichael*, 526 US 137, 148 (1999) (“[I]t would prove difficult, if not impossible, for judges to administer evidentiary rules under which a gate-keeping obligation depended upon a distinction between ‘scientific’ knowledge and ‘technical’ or ‘other specialized’ knowledge. There is no clear line that divides the one from the others.”).

not only when proffered experts should be allowed to testify about general research findings but also when they should be limited to such testimony or instead be permitted to testify regarding whether the present case is an instance of the general phenomenon. In order to accomplish this objective, we first describe the legal and scientific considerations underlying the G2i issue in greater detail, then examine how courts today deal with the issue, and finally set forth criteria by which courts might handle G2i inferences in a more informed and coherent way than they do currently.

The major contention of this Article is that expert testimony focused on 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). The analysis of whether expert testimony is sufficiently reliable to be admissible should vary depending on whether the testimony simply reports scientific findings or instead seeks to address the case at hand. Until courts recognize this point and grapple with it, they will not be able to make optimal use of science’s insights.

I. THE G2i CHALLENGE

The basic challenge inherent in reasoning from group data to individual events is not unique to the courtroom. Indeed, since all applied science depends on group data to establish what is known about a phenomenon of interest, G2i issues exist whenever group data are applied to individual cases. A multitude of fields outside the courtroom could serve as useful illustrations. For example, in medicine there is substantial research regarding the age at which women should begin having annual mammograms to screen for breast cancer. These data provide an empirical framework that doctors and patients can use to make individual decisions and help manage the risks of breast cancer and its treatment. Although the mammogram research cannot provide certainty or determine categorically what should be done in particular cases, it can and often does inform (“frame”) those decisions. Similarly, meteorologists use group data to model the trajectory and severity of storms. These models help inform policy makers who must decide whether to evacuate a community in response to the threat of a particular storm. The models provide a statistical projection of how similarly situated

storms typically evolve. Whether a particular storm is an instance of those storms previously studied, and on which the model depends, is a probabilistic estimation whose value depends on the variables studied in the original research and the methods used to study them. Further examples could be gleaned from virtually any other area of applied science.

In the courtroom, the challenges of G2i are similarly presented when expert testimony is proffered, and judges must determine daily how to handle them. For example, imagine two experts, A and B, both of whom want to testify in a trial in which one of the issues is whether the defendant doctor was negligent in failing to recommend a mammogram for the plaintiff, who now has breast cancer. Expert A will testify for the defense that research shows the optimum age for a woman to get a mammogram is between forty and fifty (the plaintiff is now thirty-eight). Expert B will testify for the plaintiff that he is aware of the research, but that in his experience, which includes seeing hundreds of female patients, some of them with histories similar to the plaintiff's, the plaintiff should have been told to get a mammogram at age thirty-five. Should both, neither, or only one of the experts be allowed to testify? In making this decision should the court consider the same types of factors, with the same threshold inquiries, or is a different analysis in order? This Article sets forth a set of best-practice guidelines that can be used to answer these types of questions.

A. The Legal Structure of G2i—Framework Evidence and Diagnostic Evidence

Since it has occasion to use just about every kind of applied science, the law constantly confronts the G2i issue. Ordinarily, in criminal and civil trials, science is introduced through expert testimony and is thought to apply to a factual issue in dispute.⁵ But, properly understood, neither evidence about the results of scientific research nor its application to the case at hand is a conventional trial fact.

Most facts disputed at trial are what Professor Kenneth Culp Davis called "adjudicative facts."⁶ According to Davis,

⁵ Issues surrounding G2i are similarly presented in administrative hearings, and much of the discussion below could be applied to them. We limit our examination to the criminal and civil-trial arenas, however, because evidentiary rules often apply differently in administrative hearings than they do in criminal and civil adjudication.

⁶ Kenneth Culp Davis, *Administrative Law Text* § 7.03 at 160 (West 3d ed 1972).

"[a]djudicative facts usually answer the questions of who did what, where, when, how, why, with what motive or intent."⁷ Davis contrasted adjudicative facts to what he called "legislative facts." Legislative facts are those facts that have relevance to legal reasoning and the fashioning of legal rules.⁸ Davis explained that "[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."⁹ In Davis's original formulation, jurors primarily found adjudicative facts and judges found legislative facts.¹⁰

Davis's adjudicative-fact/legislative-fact dichotomy is useful as a starting point, but it does not adequately capture the G2i issue as it is presented in ordinary adjudication. The general component of G2i is not "legislative" in any ordinary sense, but rather is a fact that both transcends the dispute and is relevant to adjudicating that dispute. At the same time, the individual component of G2i is not "adjudicative" in the ordinary sense, since its proof depends on the evidence for the general phenomenon.

In a series of three articles, one of us (J.M.) and Professor Laurens Walker proposed a better way to understand trial-level fact finding and, concomitantly, a better way to visualize the G2i issue in law.¹¹ They identified a new classification of facts—a hybrid between adjudicative and legislative facts—that they referred to as "social frameworks."¹² Monahan and Walker pointed out that social frameworks—or what we refer to here as "empirical frameworks"¹³—are a common form of fact-finding in the courts. As they explained, a fundamental characteristic of much scientific research is that its relevance has to be understood at two levels of generality, levels that are analytically separate. On

⁷ Id.

⁸ Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 Harv L Rev 364, 424 (1942).

⁹ Davis, *Administrative Law Text* § 7.03 at 160 (cited in note 6).

¹⁰ Id.

¹¹ See John Monahan and Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U Pa L Rev 477, 488 (1986); Laurens Walker and John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 Va L Rev 559, 570 (1987); Laurens Walker and John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 Cal L Rev 877, 879 (1988).

¹² Walker and Monahan, 73 Va L Rev at 559 (cited in note 11).

¹³ 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, and so we use the more generic term "empirical framework."

the one hand, science 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.¹⁴

The law is replete with examples of this dual use of science. Consider, for instance, cases in which the court must assess a criminal offender's likelihood of recidivism, a drug company's liability for a deformed baby, or the credibility of a witness. Jurors trying to assess these claims might want to know whether mental illness is a risk factor for violence, Bendectin causes birth defects, and polygraphs validly detect deception. Answers to each of these questions require resort to an empirical framework that is built on a foundation of research on groups. Expert explanations of these empirical frameworks may help jurors understand whether people with mental illness are more likely to be violent, whether Bendectin increases the risk for birth defects, and how good polygraphs are at detecting deception. Ultimately, however, jurors also need to determine whether, in the case before them, the criminal offender will recidivate, the plaintiff's birth defects were caused by Bendectin, and the witness is telling the truth. Experts, therefore, might also be called to help determine whether the case at hand is an instance of the empirical framework.

We refer to the scientist who testifies to the empirical framework of which a particular case might be an instance as a "framework expert." Framework experts testify to the "G" in G2i. We refer to witnesses who apply scientific data to individual cases as "diagnostic experts," because diagnostic means "serving to distinguish, identify, or determine"¹⁵ or "being a precise indication."¹⁶ Diagnostic experts testify to the "i" in G2i.¹⁷ Expert A in the mammogram case described above is a framework expert. Expert B in that case is a diagnostic expert.

¹⁴ Walker and Monahan, 76 Cal L Rev at 885 (cited in note 11).

¹⁵ *Webster's Third New International Dictionary of the English Language Unabridged* 622 (Merriam-Webster 2002).

¹⁶ *Diagnostic*, Dictionary.com, online at <http://dictionary.reference.com/browse/diagnostic?s=t> (visited May 21, 2014). Thus, our use of the term "diagnostic" is broader than the more common use of the term in medicine and psychiatry.

¹⁷ Of course, one expert could fulfill both roles.

B. Managing G2i

The operative question for courts is how to best manage the G2i issue. As Monahan and Walker explained in their development of the framework concept, the decision whether to admit expert testimony regarding the empirical framework is separate from the decision whether to admit expert testimony offering an opinion that a particular case is an instance of that empirical framework.¹⁸ The most important difference in this calculus is that diagnostic testimony cannot be admissible unless the relevant framework is also admissible, whereas the converse is not true. Even if framework evidence is admissible, extrapolation from it to the individual case may not be scientifically or legally justifiable.¹⁹

Consider expert testimony about the foibles of eyewitness identifications. Researchers have identified a number of factors that interfere with eyewitness accuracy, with some affecting the initial identification, others the memory of the identification, and still others the elicitation of the identification from its storage in memory. Examples of the first type of factor include weapon focus and cross-racial identifications, examples of the second include confounding information received between identification and recall, and examples of the third include the way lineups are presented.²⁰ This body of research is voluminous and robust, sufficiently so to be admissible in the appropriate case. But, as we describe in more detail below, almost all researchers in the field agree that the research does not permit an expert to

¹⁸ In their early work on “social frameworks,” Walker and Monahan proposed that frameworks be communicated to the jury via instructions. Walker and Monahan, 73 Va L Rev at 592–97 (cited in note 11). More recently, they have suggested additional procedures by which frameworks could be communicated to juries to include expert testimony. See John Monahan, Laurens Walker, and Gregory Mitchell, *Contextual Evidence of Gender Discrimination: The Ascendance of “Social Frameworks,”* 94 Va L Rev 1715, 1733–37 (2008). While some courts use instructions to communicate frameworks to juries—see *State v Henderson*, 27 A3d 872, 928 (NJ 2011)—frameworks are usually communicated via the testimony of expert witnesses.

¹⁹ See David L. Faigman, *Making the Law Safe for Science: A Proposed Rule for the Admission of Expert Testimony*, 35 Washburn L J 401, 414 (1996) (“[A]llowing experts to . . . apply the science to the case without research supporting their ability to do so invites expert speculation borne out of expert witness fees or a zeal for the party’s cause.”).

²⁰ For reviews of scientific studies on eyewitness testimony, see Elizabeth F. Loftus, *Eyewitness Testimony* 35–36, 52–109, 136–42 (Harvard 1979); Gary L. Wells and Elizabeth A. Olson, *Eyewitness Testimony*, 54 Ann Rev Psychology 277, 285–89 (2003); Gary L. Wells and Elizabeth F. Loftus, *Eyewitness Memory for People and Events*, in Irving B. Weiner and Randy K. Otto, eds, 11 *Handbook of Psychology: Forensic Psychology* 617, 620–27 (Wiley 2013).

offer an opinion that any particular identification is or is not accurate—whether that opinion is communicated in a categorical format (for example, “the witness made an inaccurate identification”) or a probabilistic format (“the likelihood that the witness made an accurate identification is 27 percent”); thus, these scientists conclude, the research cannot be used diagnostically to determine whether a particular eyewitness identification was accurate. If this view is correct, experts should be restricted to educating the jury about the limits of eyewitness identification.²¹

Courts are thus confronted with two fundamental options with respect to the G2i question. They can limit experts to testifying about the relevant framework or they can permit expert opinions explaining both the empirical framework and its application to the instant case. Under the first approach experts describe general background information, and jurors are left to apply it—or not—to the respective case at hand. The second approach permits expert testimony at both the framework level and its application to the particular case. Choosing between these options is the core G2i issue. Unfortunately, the courts do not always recognize that the issue even exists, much less analyze it properly.

II. THE COURTS AND G2i

Most courts have no general conception of how to manage framework and diagnostic testimony. As a practical matter, however, they must regularly decide how case specific to allow an expert to get. This Part samples the case law in this area and, more particularly, the statements that courts have made regarding the scope of proffered expert testimony. This sampling is not meant to be representative, but instead provides a tableau of cases that have considered whether experts should be limited to framework expert testimony or instead permitted (or required) to provide diagnostic expert testimony. Because these decisions ordinarily arise out of courts’ interpretation of *Frye v United States*,²² *Daubert v Merrell Dow Pharmaceuti-*

²¹ The Advisory Committee’s Note to Federal Rule of Evidence 702 explicitly contemplates this educative function, stating that “[t]he rule [] recognizes that an expert on the stand may give a dissertation or exposition of scientific or other principles relevant to the case, leaving the trier of fact to apply them to the facts.” FRE 702, Advisory Committee’s Note to the 2012 Proposed Rules.

²² 293 F 1013 (DC Cir 1923).

cal^s,²³ and their progeny, we begin with a brief review of those cases.

A. Basic Legal Standards: *Frye* and *Daubert*

Volumes have been written on *Frye* and *Daubert*, both individually and in comparison to one another.²⁴ It is generally agreed that the two tests differ substantially in their approaches to expert testimony, but nonetheless share certain commonalities. This Section does not rehearse these points in detail but rather discusses *Frye* and *Daubert* in terms of how they might affect G2i considerations.

1. *Frye*.

The basic requirement of *Frye* is that the “thing” upon which expert evidence is based must have “gained general acceptance in the particular field in which it belongs.”²⁵ This language is famously ambiguous. Among other issues, it leaves unclear precisely what must be generally accepted (the general methodology behind the expert’s testimony, the way in which the expert applied the methodology, or both); how to define the relevant field; and how many in the relevant field must accept the methodology for it to be accepted.²⁶

Thus, for instance, if the expert wants to provide framework evidence about eyewitness testimony, *Frye* leaves up in the air whether the court needs to examine not just the acceptability of eyewitness research generally but also the precise findings the expert will present. It also leaves unclear whether the relevant field to be consulted consists of eyewitness researchers who have done the type of research the expert will describe, all eyewitness researchers, or all cognitive psychologists, and, if there is disagreement about research results (which there often is), the extent to which that matters. Similarly, if the expert wants to offer

²³ 509 US 579 (1993).

²⁴ Numerous symposia collect articles from a number of authors on the rules that govern the admissibility of expert testimony and how they should be interpreted. See generally, for example, Symposium, *The Daubert Hearing: From All the Critical Perspectives*, 46 UC Davis L Rev 737 (2013); Symposium, *Expertise in the Courtroom: Scientists and Wizards*, 52 Vill L Rev 679 (2007); Symposium, *Expert Admissibility Symposium: What Is the Question? What Is the Answer? How Should the Court Frame a Question to Which Standards of Reliability Are to Be Applied?*, 34 Seton Hall L Rev 1 (2003).

²⁵ *Frye*, 293 F at 1014.

²⁶ See Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States a Half-Century Later*, 80 Colum L Rev 1197, 1215–16 (1980).

diagnostic testimony about the causes of an illness, *Frye* leaves unanswered whether the court need only evaluate the acceptability of the general causative theory espoused by the expert or also must look at the acceptability of the expert's precise evaluation method and reasoning. Nor does it tell the court what types of doctors and how many of them must subscribe to the expert's practices.

Frye is not necessarily incompatible with an empirically sophisticated use of expert evidence, but the test does not promote it either. In contrast, *Daubert's* focus on the methods and principles underlying proffered expertise has exactly this effect.

2. *Daubert*.

The basic holding of *Daubert*, which construed Federal Rule of Evidence 702's definition of admissible expertise, has three components, all of which are important to G2i analysis. First, the expert evidence must be relevant, that is, it must relate to an issue in the case.²⁷ The Court referred to this element as one of "fit," in that the empirical basis for the evidence must address a fact in dispute.²⁸ Second, the expert must be qualified to testify on the subject at hand. This element is fairly permissive and can be met "by knowledge, skill, experience, training, or education."²⁹ These first two requirements can be read into the *Frye* test as well, or at least added onto it as friendly amendments. The key distinction between *Frye* on the one hand and *Daubert* on the other is *Daubert's* third component, which states that the "[p]roposed testimony must be supported by appropriate validation," what *Daubert* referred to as "good grounds."³⁰ According to the Court, "the requirement that an expert's testimony pertain to 'scientific knowledge' establishes a standard of evidentiary reliability."³¹

This third prong of *Daubert* represents a fundamental change from *Frye*. Whereas *Frye* speaks of general acceptance, *Daubert's* focus is validity. More specifically, *Daubert* mandated

²⁷ See *Daubert*, 509 US at 587.

²⁸ *Id.* at 591.

²⁹ FRE 702.

³⁰ *Daubert*, 509 US at 590.

³¹ *Id.* The Court likened "evidentiary reliability"—which it deemed to be the legally pertinent inquiry of "trustworthiness"—to what scientists refer to as "validity." *Id.* at 590–91 n 9. The Court clearly stated that "[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*." *Id.* In this Article, we use the scientist's preferred term of "validity."

that under the Federal Rules of Evidence trial courts have the responsibility to examine the methodologies and principles underlying proffered expert testimony to determine whether those principles and methods are sufficiently valid.³² Under Rule 104(a), a court must find this preliminary fact by a preponderance of the evidence.³³

Daubert also suggested that, in exercising their gatekeeping function of evaluating the underlying basis for proffered scientific evidence, judges consider four factors: testability, error rates and adequacy of standards, peer review and publication, and general acceptance.³⁴ According to the Court, the first factor is important because the scientific status of testimony is tied to “its falsifiability, or refutability, or testability.”³⁵ Thus, “a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (*and has been*) tested.”³⁶ The second factor the Court identified—the “error rate” associated with the “particular scientific technique”³⁷—was not discussed in detail, but the Court stated that courts should “ordinarily” also consider this factor, along with the “standards” that exist to control “the technique’s operation.”³⁸ The third factor the Court noted was whether the research “has been subjected to peer review and publication.”³⁹ The Court emphasized that this factor was “not a *sine qua non* of admissibility,” but was a “relevant . . . consideration in assessing . . . scientific validity.”⁴⁰ Finally, the Court stated, “‘general acceptance’ can yet have a bearing on the

³² *Id.* at 592–93 (stating that the trial judge must make “a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue”).

³³ See FRE 104(a); *Bourjaily v United States*, 483 US 171, 175–76 (1987) (holding that determinations of fact under FRE 104(a) must be proved by a preponderance of the evidence).

³⁴ As we emphasize in the Conclusion, the four “*Daubert* factors” are merely guidelines to help courts assess expert testimony. At the same time, as Justice Scalia pointed out concurring in *Kumho Tire Company v Carmichael*, although “the *Daubert* factors are not holy writ, in a particular case the failure to apply one or another of them may be unreasonable, and hence an abuse of discretion.” *Kumho Tire Co v Carmichael*, 526 US 137, 159 (1999) (Scalia concurring).

³⁵ *Daubert*, 509 US at 593, quoting Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* 37 (Routledge 5th ed 1989).

³⁶ *Daubert*, 509 US at 593 (emphasis added).

³⁷ *Id.* at 594.

³⁸ *Id.*

³⁹ *Id.* at 593.

⁴⁰ *Daubert*, 509 US at 593–94.

inquiry.”⁴¹ Lack of “[w]idespread acceptance,” the Court observed, may lead judges to view the particular evidence “with skepticism.”⁴²

Two other Supreme Court decisions construing *Daubert* need to be mentioned in order to round out the evidentiary picture. The first is *General Electric Co v Joiner*,⁴³ decided four years later. While *Joiner* is probably best known for its holding that appellate courts owe deference to a trial court’s admissibility rulings,⁴⁴ of most relevance to the G2i issue was *Joiner*’s clarification of the language in *Daubert* stating that, in carrying out its gatekeeper role, a trial court’s “focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.”⁴⁵ One could conclude that this language was meant to limit the reach of the district court’s responsibilities. But the *Joiner* Court observed that “nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert.”⁴⁶ This statement suggests that a court’s gatekeeping obligations extend to all components of an expert’s proposed testimony, from the most general to the most specific.⁴⁷

After *Joiner*, the big question involving expert evidence concerned whether *Daubert*’s new rules applied only to “scientific” evidence or whether they applied as well to “technical and specialized knowledge,” the two other types of expertise mentioned in Federal Rule of Evidence 702. Several lower courts held that *Daubert* applied only to expert testimony characterized as “scientific.”⁴⁸ In *Kumho Tire Co v Carmichael*,⁴⁹ the Court rejected the science/nonscience distinction.⁵⁰ Justice Stephen Breyer, writing for the Court, said that “it would prove difficult, if not impossible, for judges to administer evidentiary rules under which a gatekeeping obligation depended upon a distinction be-

⁴¹ *Id.*

⁴² *Id.*

⁴³ 522 US 136 (1997).

⁴⁴ *Id.* at 139.

⁴⁵ *Daubert*, 509 US at 595.

⁴⁶ *Joiner*, 522 US at 146.

⁴⁷ Rule 702, amended in 2000, expressly requires, among other things, that expert testimony be “based on sufficient facts or data,” and that “the expert has reliably applied the principles and methods to the facts of the case.” FRE 702(b), (d).

⁴⁸ See, for example, *Iacobelli Construction v County of Monroe*, 32 F3d 19, 25 (2d Cir 1994); *Thornton v Caterpillar, Inc.*, 951 F Supp 575, 577 (D SC 1997); *United States v Starzeczpyzel*, 880 F Supp 1027, 1041 (SDNY 1995).

⁴⁹ 526 US 137 (1999).

⁵⁰ *Id.* at 148.

tween ‘scientific’ knowledge and ‘technical’ or ‘other specialized’ knowledge. There is no clear line that divides the one from the others.”⁵¹ Moreover, the Court did not accept the proposition that jurors would be more impressed by “scientific” testimony than they might be by other categories of expert evidence. Instead, the Court stated, all expert testimony has the potential to overawe jurors.⁵²

In language that makes no distinction between framework and diagnostic testimony, *Kumho Tire* emphasized that *Daubert* “imposes a special obligation upon a trial judge to ‘ensure that any and all scientific testimony . . . is not only relevant, but reliable,’”⁵³ an obligation that includes “mak[ing] certain that an expert . . . employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.”⁵⁴ The Court extended this special obligation to any “knowledge [that] might become the subject of expert testimony,” not just of the scientific variety.⁵⁵ The Court stated emphatically: “And where such testimony’s factual basis, data, principles, methods, or their application are called sufficiently into question . . . the trial judge must determine whether the testimony has ‘a reliable basis in the knowledge and experience of [the relevant] discipline.’”⁵⁶

Daubert and Rule 702 thus endeavor to empower judges to screen proffered expert evidence of whatever variety and proffered for whatever purpose. Although debate continues regarding just how rigorous these gatekeeping requirements are, or should be, there is little question that *Daubert* and its progeny changed the terms of the debate.⁵⁷ As the next Section describes, in many respects the G2i issue—little noted before *Daubert* was decided in 1993—is both a component and a consequence of *Daubert*.

⁵¹ *Id.*

⁵² See *id.*

⁵³ *Kumho Tire*, 526 US at 147, quoting *Daubert*, 509 US at 589.

⁵⁴ *Kumho Tire*, 526 US at 152.

⁵⁵ *Id.* at 147.

⁵⁶ *Id.* at 149, quoting *Daubert*, 509 US at 592.

⁵⁷ Compare David L. Faigman, *The Daubert Revolution and the Birth of Modernity: Managing Scientific Evidence in the Age of Science*, 46 UC Davis L Rev 893, 895–97 (2013), with Edward J. Imwinkelried, *The Meaning of “Appropriate Validation” in Daubert v. Merrell Dow Pharmaceuticals, Inc., Interpreted in Light of the Broader Rationalist Tradition, Not the Narrow Scientific Tradition*, 30 Fla St U L Rev 735, 739 (2003).

B. How Courts Handle the G2i Challenge

Neither *Frye* nor *Daubert* directly addressed the level of specificity to which they permit or require experts to testify. For reasons given earlier, we think the basic choice is between framework evidence alone and framework evidence coupled with particularized proof. But, to the extent they consider it at all, courts appear to be inconsistent in how they make this choice and the reasons they give for it. Below we illustrate this point by examining court decisions in several different areas of expertise.

1. Eyewitness testimony.

The one scientific area in which courts have often focused on the divide between the general and the specific is eyewitness-identification research. The likely reason courts attend to G2i in eyewitness cases is that research psychologists who testify on this subject do so. As we noted earlier, eyewitness researchers are nearly unanimous in conceding that while they have much to offer in regard to framework testimony, they cannot validly make statements about individual cases. Two of the most prominent researchers have bluntly stated that “[t]he state of the science . . . does not permit an assessment of the accuracy of an individual eyewitness.”⁵⁸ Further, they assert, “any statement that allows the jury to infer that the expert believes a specific witness to be inaccurate, whether in response to a direct or hypothetical question, is a scientifically unsupported use of expert testimony.”⁵⁹

Consistent with this consensus among the experts, most courts assessing the admissibility of expert testimony about eyewitnesses demonstrate an appreciation that science begins—and often ends—from the perspective of studying groups. In *United States v Smith*,⁶⁰ for example, the court specifically held that the eyewitness expert was “allowed to educate the jury about the psychological literature,” but “was not permitted to discuss witnesses in this case at all.”⁶¹ *Smith* and other courts have pointed out that by limiting the expert’s testimony to factors that generally interfere with eyewitness perception and re-

⁵⁸ Brian L. Cutler and Gary L. Wells, *Expert Testimony regarding Eyewitness Identification*, in Jennifer L. Skeem, Kevin S. Douglas, and Scott O. Lilienfeld, eds, *Psychological Science in the Courtroom: Consensus and Controversy* 100, 113 (Guilford 2009).

⁵⁹ *Id.*

⁶⁰ 621 F Supp 2d 1207 (MD Ala 2009).

⁶¹ *Id.* at 1218.

call, the expert can help jurors appreciate the limitations of eyewitness identifications without trenching on the jury's role of evaluating witness credibility.⁶² As explained by the court in *United States v Hines*,⁶³ "[t]he function of the expert here is not to say to the jury—you should believe or not believe the eyewitness.' . . . All that the expert does is provide the jury with more information with which the jury can then make a more informed decision."⁶⁴

Not all courts follow the lead of the experts, however. Some courts still insist that eyewitness experts extend their assessments to the accuracy of particular witnesses. In *State v McLean*,⁶⁵ for example, the court stated that an expert's failure to offer an opinion about the reliability of the witness's identification precluded his testimony. According to the *McLean* court, the trial judge properly excludes expert testimony on eyewitness identification when the expert "did not interview the witnesses in [the] case, did not observe their trial testimony, and did not visit the crime scene."⁶⁶ A similar view was expressed by the Connecticut Supreme Court in *State v McClendon*,⁶⁷ which excluded expert testimony explaining the research relating to the accuracy of eyewitnesses on the ground that "[t]he jury [must] have the opportunity to assess the witnesses' credibility on the basis of what is presented at trial and not solely on general principles."⁶⁸ The Texas Court of Criminal Appeals has also embraced the position that eyewitness experts must apply their science as nearly as possible to the particular case.⁶⁹

These differing approaches to expert testimony about eyewitness identification are replicated in other fields of cognitive psychology.⁷⁰ They suggest that courts have yet to devise a uniform

⁶² Id at 1218–19 (“[E]ducating the jury about this research . . . is an important step along the road to using improved scientific knowledge to create more accurate and fair legal proceedings. . . . Applying this research to the facts of this case, however, is within the sole province of the jury.”).

⁶³ 55 F Supp 2d 62 (D Mass 1999).

⁶⁴ Id at 72.

⁶⁵ 645 SE2d 162 (NC App 2007).

⁶⁶ Id at 165.

⁶⁷ 730 A2d 1107 (Conn 1999).

⁶⁸ Id at 1116.

⁶⁹ See, for example, *Williams v State*, 895 SW2d 363, 366–67 (Tex Crim App 1994).

⁷⁰ Compare *United States v Dixon*, 261 Fed Appx 800, 802, 804–05 (5th Cir 2008) (holding that the district court did not abuse its discretion in excluding the defense expert's proffered testimony concerning the general phenomenon of false confessions, noting that the witness had not examined the defendant and “was not prepared to opine on whether [his] confessions were false”), with *United States v Jacques*, 784 F Supp 2d 59,

approach to G2i issues even when the experts are relatively clear about the nature of their expertise. The same mixed message is found in the courts' treatment of other areas of expert testimony.

2. Medical and psychological conditions.

In contrast to eyewitness experts, the professions of medicine (including psychiatry) and clinical psychology have long practiced particularization in ordinary practice. Although their disciplines are based on general knowledge, both professions need to make and act upon—that is, treat—individuals with particular diagnoses. For the most part, courts have, without a second thought, allowed these experts to provide the same service in the courtroom. Thus, for instance, medical experts, prodded by their attorneys, often state that they can proffer an opinion about a particular individual with “a reasonable degree of medical certainty,” and courts are willing to entertain conclusions framed in those terms.⁷¹ Psychiatrists and clinical psychologists are usually permitted to make the same types of assertions, and thus often testify, for instance, that within a reasonable degree of psychological certainty, a particular defendant has schizophrenia and, as a result, was unable to appreciate the wrongfulness of his conduct at the time of the offense.⁷²

Judicial resistance to this practice does occur occasionally, however. For instance, in the context of testimony by mental-health professionals, courts have been leery of opinions that rely on a relatively novel syndrome or theory, such as rape-trauma syndrome or battered woman syndrome. In such cases, whether applying *Frye* or *Daubert*, courts have either excluded the testimony entirely or permitted only framework evidence, without allowing application to the specific case.⁷³

63 (D Mass 2011) (reasoning that an expert's diagnostic testimony was inadmissible, as it would amount to an improper statement regarding guilt).

⁷¹ Jeff L. Lewin, *The Genesis and Evolution of Legal Uncertainty about “Reasonable Medical Certainty,”* 57 Md L Rev 380, 397–406 (1998).

⁷² See, for example, *United States v Grigsby*, 712 F3d 964, 968 (6th Cir 2013) (describing testimony by a mental-health professional that there was a “reasonable psychological certainty” that the defendant had a mental illness at the time of the offense).

⁷³ See, for example, *State v Saldana*, 324 NW2d 227, 230 (Minn 1982) (excluding testimony about rape-trauma syndrome on the ground that “[t]he jury must not decide this case on the basis of how most people react to rape or on whether [the victim's] reactions were the typical reactions of a person who has been a victim of rape. Rather, the jury must decide what happened in *this case*.”); *Commonwealth v Goetzendanner*, 679 NE2d 240, 243–46 (Mass App 1997) (ruling that expert testimony about domestic vio-

Courts have also been somewhat more cautious about diagnostic testimony in “toxic tort” cases involving the causes of certain types of medical conditions, such as cancer and birth defects. In this particular area of medical testimony, courts explicitly recognize the G2i issue by distinguishing between evidence of “general causation” (framework evidence) and “specific causation” (diagnostic evidence) and requiring proof of the first before permitting the second.⁷⁴ Consider, for example, a claim that the plaintiff’s leukemia was caused by exposure to perchloroethylene (PCE), a chemical in dry cleaning fluid.⁷⁵ Virtually all courts today require credible scientific proof that PCE causes leukemia in populations exposed to it at dosages comparable to the dosage experienced by the plaintiff. Only if this general proof of causation is met do courts permit⁷⁶ or require⁷⁷ the plaintiff to provide additional expert evidence of specific causation, to the effect that his or her own leukemia was caused by PCE. Typically in these cases, one or more experts will testify to the research framework and a different set of experts will testify that the plaintiff’s case is an instance of that framework.⁷⁸

Because it focuses courts’ attention on the methodologies underlying expert opinions, *Daubert* has probably spurred the courts’ willingness to distinguish between these different levels of causation in toxic tort cases. In fact, a computer search we conducted revealed that whereas only seventeen such cases made the distinction prior to *Daubert*, many hundreds did so after *Daubert*.⁷⁹ Unfortunately, however, these courts usually fail

lence and battered woman’s syndrome in general terms was properly admitted to explain victim’s conduct, so long as the expert did not offer an opinion or diagnosis that victim suffers from syndrome); *People v Christel*, 537 NW2d 194, 201 (Mich 1995) (ruling that an expert may explain generalities of battered woman’s syndrome to describe the uniqueness of particular behavior at issue, but an expert may not offer an opinion whether the victim was a battered woman).

⁷⁴ See *In re Rezulin Products Liability Litigation*, 369 F Supp 2d 398, 401–02 (SDNY 2005) (collecting cases).

⁷⁵ See, for example, *Magistrini v One Hour Martinizing Dry Cleaning*, 180 F Supp 2d 584, 589 (D NJ 2002).

⁷⁶ *In re Rezulin Products Liability Litigation*, 369 F Supp 2d at 438.

⁷⁷ See, for example, *In re Aredia and Zometa Products Liability Litigation*, 483 Fed Appx 182, 191 (6th Cir 2012) (“Because Plaintiff failed to demonstrate an essential element of her case, specific causation, the grant of summary judgment was appropriate.”).

⁷⁸ See, for example, *Pluck v BP Oil Pipeline Co*, 640 F3d 671, 675 (6th Cir 2011) (in which the plaintiffs hired two experts on the general link between contaminants and their injuries, only one of whom testified to specific causation).

⁷⁹ Since *Daubert* interpreted Federal Rule of Evidence 702, the computer search used Westlaw’s “Allfeds” database, using the following search terms: “general /10 causation and specific /10 causation.” This search was run for all cases decided prior to *Daub-*

to consider carefully how *Daubert* should apply to specific causation testimony; rather, once general causation is proven, they admit testimony on specific causation relying on the same liberal grounds seen in more routine medical and psychological cases.⁸⁰

3. Employment discrimination.

In employment-discrimination litigation involving gender bias in hiring and promotion, the G2i issue has been highly contentious.⁸¹ In *Wal-Mart Stores, Inc v Dukes*,⁸² for example, Dr. William Bielby, a research sociologist testifying for the plaintiffs, described general research findings on the prevalence of gender stereotypes in the workplace and the conditions under which employers are more or less likely to rely on these gender stereotypes.⁸³ After providing this framework testimony, Dr. Bielby went on to explicitly link these general research results to the specific defendant in the case, Wal-Mart, offering the diagnostic opinions: (1) that “[s]ubjective and discretionary features of the company’s personnel policy and practice make decisions about compensation and promotion vulnerable to gender bias,” (2) that “there are significant deficiencies in the company’s

ert (June 28, 1993) and for all cases decided after *Daubert*. The pre-*Daubert* search identified 29 cases in which these terms were present. Of these cases, 17 involved substantive discussion of the principles of general and specific causation. The post-*Daubert* search identified 797 cases in which these terms were present, with subsequent review of these cases indicating that 512 dealt with the subject substantively. We further tested the effect of *Daubert* by searching the frequency with which courts used the term “differential etiology,” which is the predominant methodology—at least as recognized by courts today—for reasoning from general causation to specific causation. In brief, differential etiology is the method by which one putative cause is ruled in via general causation while other putative causes are sought to be ruled out. The search for this term prior to *Daubert* found no federal cases using it. Following *Daubert*, 116 reported decisions used this term. An unpublished memorandum reporting this survey is available from the authors.

⁸⁰ See Joseph Sanders, *Applying Daubert Inconsistently? Proof of Individual Causation in Toxic Tort and Forensic Cases*, 75 Brooklyn L Rev 1367, 1374, 1403–04 (2010) (analogizing the admissibility standards applied to specific causation testimony in toxic tort cases to the “liberal” standards applied to forensic testimony in criminal cases and stating that “this liberal approach unnecessarily helps to perpetuate the status quo”).

⁸¹ Compare Monahan, Walker, and Mitchell, 94 Va L Rev at 1734 (cited in note 18), and David L. Faigman, Nilanjana Dasgupta, and Cecilia L. Ridgeway, *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 Hastings L J 1389, 1431 (2008), with Melissa Hart and Paul M. Secunda, *A Matter of Context: Social Framework Evidence in Employment Discrimination Class Actions*, 78 Fordham L Rev 37, 59–66 (2009), and Susan T. Fiske and Eugene Borgida, *Best Practices: How to Evaluate Psychological Science for Use by Organizations*, 31 Rsrch in Organizational Behav 253, 269–70 (2011).

⁸² 131 S Ct 2541 (2011).

⁸³ See id at 2549.

policies and practices for identifying and eliminating barriers to equal employment opportunity at Wal-Mart,” and (3) that these deficient policies and practices “contribute[d] to disparities between men and women in their compensation and career trajectories at the company.”⁸⁴

While the Ninth Circuit approved of such diagnostic testimony by research scientists in gender-discrimination class actions,⁸⁵ the US Supreme Court did not:

The only evidence of a “general policy of discrimination” respondents produced was the testimony of Dr. William Bielby, their sociological expert. Relying on “social framework” analysis, Bielby testified that Wal-Mart has a “strong corporate culture,” that makes it “vulnerable” to “gender bias.” He could not, however, “determine with any specificity how regularly stereotypes play a meaningful role in employment decisions at Wal-Mart. . . . [Bielby’s expert testimony] is worlds away from “significant proof” that Wal-Mart “operated under a general policy of discrimination.”⁸⁶

The Court went on to comment directly upon the difference between expert testimony on general frameworks and expert testimony on case-specific facts:

Bielby’s conclusions in this case have elicited criticism from the very scholars on whose conclusions he relies for his social-framework analysis. See Monahan, Walker, and Mitchell, Contextual Evidence of Gender Discrimination: The Ascendance of “Social Frameworks,” 94 Va. L. Rev. 1715, 1747 (2008) . . . (“[A] social framework necessarily contains only general statements about reliable patterns of relations among variables . . . and goes no further. . . . Dr. Bielby claimed to present a social framework, but he testified about social facts specific to Wal-Mart”); *id.*, at 1747–1748 (“Dr. Bielby’s report provides no verifiable method for measuring and testing any of the variables that were crucial to his conclusions and reflects nothing more than Dr. Bielby’s ‘expert judgment’ about how general stereotyping research applied to

⁸⁴ Declaration of William T. Bielby, PhD in Support of Plaintiffs’ Motion for Class Certification, *Dukes v Wal-Mart Stores, Inc.*, Civil Action No C-01-2252, *5–6, 41 (ND Cal filed 2003).

⁸⁵ *Dukes v Wal-Mart, Inc.*, 509 F3d 1168, 1179 (9th Cir 2007).

⁸⁶ *Wal-Mart Stores*, 131 S Ct at 2553–54 (citations omitted).

all managers across all of Wal-Mart's stores nationwide for the multi-year class period").⁸⁷

It remains too soon to tell whether *Wal-Mart* will sharpen the distinction between general framework testimony and case-specific diagnostic testimony in other areas of the law. But the case demonstrates that the Supreme Court is at least cognizant of the issue.

4. Forensic identification.

In contrast to the incipient judicial recognition of the G2i issue in cases involving testimony about eyewitnesses, psychological and medical conditions, and employment discrimination, courts have usually ignored the issue in cases involving use of the physical sciences to identify a perpetrator or victim. Thus, courts do not question diagnostic testimony from forensic-identification experts on firearms, fingerprints, and bite marks. Indeed, such experts very often are allowed to testify that there is a match "to the exclusion of all other [guns, fingers, or teeth] in the world."⁸⁸ In contrast, DNA-testing experts, who employ the most powerful scientific technology available for evaluating forensic trace evidence, do not offer conclusions regarding whether the DNA found at the scene of the crime can be identified as the defendant's DNA. Instead, DNA experts, and the courts following their lead, insist on general probability statements regarding the likelihood of finding the "match" randomly in the population.⁸⁹

As these examples illustrate, in identification cases courts largely adopt, at least as their default position, the respective professions' own claims regarding their practitioners' ability to make particularized judgments. Hence, despite the fact that DNA testing has vastly greater power as a forensic technology than any other forensic technology now existing, results about an individual's DNA are expressed with reference to group data,

⁸⁷ Id at 2553–54 n 8.

⁸⁸ Brian J. Heard, *Forensic Ballistics in Court: Interpretation and Presentation of Firearms Evidence* xvii (John Wiley & Sons 2013). See also, for example, *United States v Allen*, 207 F Supp 2d 856, 868 (ND Ind 2002) (footwear identification); *United States v Havard*, 117 F Supp 2d 848, 852 (SD Ind 2000) (latent fingerprints). See also generally Michael J. Saks and Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 Vand L Rev 199 (2008).

⁸⁹ See, for example, *People v Nelson*, 185 P3d 49, 64 (Cal 2008) (discussing debates over various statistical methods of calculating DNA-match probabilities but assuming that testimony will be formulated on some sort of probability estimate).

whereas results from many of the other forensic-identification sciences are presented with particularity and categorically or without any specific probability estimates.⁹⁰ Courts do not even remark upon this differential treatment, much less recognize that it might be inappropriate.

5. Summary.

This extremely brief survey of judicial treatment of the G2i issue in a sampling of areas reveals several themes. First, neither of the foundational cases—*Frye* and *Daubert*—directly addresses the issue, although *Daubert's* holding, along with *Joiner* and *Kumho Tire*, does provide incentive to investigate the validity of testimony that goes beyond framework evidence. Second, courts often do not seem to recognize that every case involving expert testimony involves a choice between allowing both framework and diagnostic testimony and allowing only framework testimony; further, even when they are aware of this choice, as in toxic tort cases, they do not carefully consider the evidentiary standards that should apply, especially as regards diagnostic testimony. Third, whether they recognize the G2i issue or not, courts tend to follow the practice the experts themselves have adopted, at least when a consensus among them exists; thus, in most cases, the expert field, not the judiciary, appears to be dictating whether framework or framework-plus-diagnostic testimony is allowed.

This deference to professional practice is more in line with *Frye* than with *Daubert*. *Frye* merely required that the proffered expert opinion be based on principles or practices that are generally accepted in the field from which it comes.⁹¹ If professions such as medicine, clinical psychology, and the forensic-identification sciences particularize in practice, then their particularizing would seem to pass muster under *Frye*, at least if their professional practice is relevant to (or fits) the legally relevant issue in dispute. *Daubert*, in contrast, does not place deference to professional fields at the center of the evidentiary analysis, but instead calls upon judges to independently assess the methods and principles underlying the proffered opinion in order

⁹⁰ See National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* 141 (National Academy 2009); Jennifer E. Laurin, *Remapping the Path Forward: Toward a Systemic View of Forensic Science Reform and Oversight*, 91 *Tex L Rev* 1051, 1060 (2013).

⁹¹ *Frye*, 293 F at 1014.

to determine its reliability. Under *Daubert*, therefore, simply because a field claims the ability to apply general research to particular cases does not make it so. Courts, as gatekeepers, are expected to determine whether it is so.

The remainder of this Article assumes that courts should independently assess whether G2i testimony should occur. On that assumption, we propose a set of best-practice guidelines that courts should take into account in addressing the admissibility of both framework testimony and diagnostic testimony.

III. ADMISSIBILITY CRITERIA FOR FRAMEWORK AND DIAGNOSTIC EVIDENCE

The courts have not developed a consistent approach to the G2i problem. However, a perusal of *Daubert* and *Frye* and the cases construing them reveals five criteria that are consistently identified as necessary considerations in determining the admissibility of scientific expert testimony: (1) relevance, (2) qualifications, (3) scientific validity, (4) added value (or helpfulness), and (5) unfair prejudice. This Part explores how these factors apply to the challenges posed by G2i, differentiating with respect to each factor between framework and diagnostic evidence.

A. Relevance (“Fit”)

As used here, relevance concerns what was called “materiality” at common law and what *Daubert* subsequently referred to as “fit.”⁹² Relevance, in the sense of fit, is part of but also distinguishable from the inquiry into whether expert evidence is helpful.⁹³ Expert evidence that is immaterial to, or does not fit, a proposition in the case cannot be helpful.⁹⁴ Rule 702 recognizes this fact by providing that expert testimony must “help the trier of fact to understand the evidence or to determine a fact in issue.”⁹⁵ However, expert testimony that relates to an issue in the case may still be unhelpful if it is within the jury’s common

⁹² *Daubert*, 509 US at 591, citing *United States v Downing*, 753 F2d 1224, 1242 (3d Cir 1985).

⁹³ *Daubert*, 509 US at 591 (observing that the language “assist the trier of fact” . . . goes primarily to relevance”).

⁹⁴ Jack B. Weinstein and Margaret A. Berger, 3 *Weinstein’s Evidence* ¶ 702[02] at 702–18 (1988) (“Expert testimony which does not relate to any issue in the case is not relevant and, ergo, non-helpful.”).

⁹⁵ FRE 702(a).

understanding. This Section examines relevance analysis as it relates to fit; the separate helpfulness criterion is examined in a later Section.

In the context of expert testimony, the fit requirement has two different meanings that are especially pertinent to the issues surrounding G2i. The first refers to the traditional question whether the expert's opinion relates to some specific issue in dispute. We refer to this as an issue of "legal fit." Legal fit calls upon courts to determine whether the expert testimony proffered in the case is material to a factual question that the substantive law requires to be answered. The second variation, one that is unique to expert evidence, is whether the research basis for the expert's opinion generalizes to the legal issues in dispute. We refer to this as an issue of "empirical fit." Empirical fit calls upon courts to determine whether the expert testimony proffered in the case is based on research methods that relate to the factual question in issue. Both of these involve the fit between proffered expertise and the legal issues in dispute, but they require separate analysis, depending upon whether the expert evidence is framework or diagnostic in orientation.

1. Framework testimony.

The case of *United States v Carmel*⁹⁶ nicely illustrates how both legal and empirical "fit" are important to a court's relevance assessment in a case involving framework testimony. In *Carmel*, the defendant claimed that his mail and wire fraud convictions were a product of his compulsive-gambling disorder.⁹⁷ In short, he asserted that his inability to control his gambling compelled him to commit mail and wire fraud and that his compulsion rendered him "insan[e]" under applicable law.⁹⁸ The basis for this claim was expert framework evidence that people with compulsive-gambling disorders lack volitional control.

In the course of upholding the exclusion of the evidence, the Seventh Circuit discussed the testimony in terms of both legal and empirical fit. The court first observed that following John Hinckley's attempt to assassinate President Ronald Reagan, Congress changed the insanity defense and removed the volitional

⁹⁶ 801 F2d 997 (7th Cir 1986).

⁹⁷ *Id.* at 998.

⁹⁸ *Id.*

prong from the test.⁹⁹ Because the “volitional prong is the only basis for an insanity defense based on a compulsive gambling disorder,” this change in the law made any proffered expert testimony on a defendant’s gambling disorder irrelevant to a claim of insanity.¹⁰⁰ This part of *Carmel* illustrates the first, and more conventional, understanding of relevance or fit—legal fit.

If evidence lacks legal fit, that should usually end the matter. But in *Carmel* it did not, because the defendant had committed his crimes prior to Congress’s amendment of the federal insanity test and the government did not claim that the change was retroactive. Thus, for the defendant, lack of volitional control continued to be a material fact in his insanity claim. Nonetheless, the court still found the expert evidence to be irrelevant, this time relying on the fit concern that is special to expert testimony—empirical fit.¹⁰¹ The empirical-fit issue concerned whether the research available on habitual gambling could be extended to the nongambling offenses with which the defendant was charged. Applying *Frye*, the *Carmel* court concluded that this proposition was not generally accepted in the relevant scientific discipline, because “there does not exist the requisite nexus between compulsive gambling and non-gambling offenses generally (and mail and wire fraud in particular) for a compulsive gambling disorder to serve as the basis of an insanity defense to such offenses.”¹⁰² With respect to this second variation of fit, the court determined that research proffered on lack of volitional control could not be generalized to nongambling offenses, whatever its applicability to gambling.¹⁰³ It thus was irrelevant to—it did not empirically fit—the issues in dispute under applicable law.

As *Carmel* demonstrates, the concept of empirical fit possesses an element of logical deduction to it, but also merges with the separate question of external validity. External validity concerns whether the results obtained in a study can be said to hold for—are generalizable to—groups or populations not specifically included in the study: in *Carmel* the research on compulsive gambling did not generalize to situations involving commission

⁹⁹ Insanity Defense Reform Act of 1984 § 402(a), Pub L No 98-473, 98 Stat 1837, 2057, codified as amended at 18 USC § 17.

¹⁰⁰ *Carmel*, 801 F2d at 998 n 1.

¹⁰¹ *Id.* at 999.

¹⁰² *Id.*

¹⁰³ *Id.*

of other offenses to feed the gambling habit.¹⁰⁴ External validity is to be distinguished from internal validity—what we later in this Article call “scientific validity”—which concerns whether the methods and principles used in the research support the results reached in the study.¹⁰⁵

Internally valid framework evidence might have weak external validity for a number of reasons. As in *Carmel*, it may simply not address the legal question. Or it may address the legal question, but not in a way that is generalizable to the real world. For instance, the relevant research might have been conducted in a laboratory setting or in a unique geographical area that only minimally replicates everyday life in the rest of the world. It might have been carried out using students rather than a sample representative of the population of interest. Or it might have been conducted so long ago that its findings are no longer pertinent.¹⁰⁶ The question of empirical fit asks whether particular scientific results have external validity with respect to the circumstances the law makes relevant.

2. Diagnostic testimony.

Relevance, or fit, must also be considered in regard to particularized expert evidence and, indeed, might play a disproportionate role there. A diagnostic expert is, by definition, addressing the individual facts of the case and thus will usually meet the legal-fit threshold. But courts must also recognize that fit may be lacking between the empirically supported basis for the diagnostician’s work outside of the courtroom and the legal needs of the courtroom.

A now-classic illustration of the possible lack of empirical fit between what medical doctors do in practice versus what they are called upon to do in court concerns the difference between what doctors call “differential diagnosis” and what the courts increasingly refer to as “differential etiology.”¹⁰⁷ In ordinary medical

¹⁰⁴ See John Monahan and Laurens Walker, eds, *Social Science in Law: Cases and Materials* 63–64 (Foundation 8th ed 2014).

¹⁰⁵ Id.

¹⁰⁶ See Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* 51–54 (Houghton Mifflin 1979) (discussing generalization across persons, settings, and time).

¹⁰⁷ For an excellent discussion and citation of authorities contrasting “differential diagnosis” and “differential etiology,” see Justice Brent Benjamin’s dissent in *San Francisco v Wendy’s International, Inc*, 656 SE2d 485, 505–08 (W Va 2007) (Benjamin dissenting).

practice, doctors are interested in, and have been trained in, diagnosing disease for the purpose of making treatment decisions.¹⁰⁸ Determining, for example, whether a person who arrives at the emergency department with chest pains has nonischemic chest pain or acute coronary ischemia is necessary for deciding proper treatment.¹⁰⁹ If the patient receives the latter diagnosis, however, most doctors do not, and are not trained to, determine its cause. Yet causation may be the key legal issue for a plaintiff who claims that her acute coronary ischemia was the result of ingesting the defendant manufacturer's drug.

The important lesson here is that courts should be clear regarding the operative fact in dispute. In the above example, is it "What is the illness?" or is it "what *caused* the illness?" More generally, courts cannot assume that what some professionals do when they determine that a particular case is an instance of an empirical framework in their ordinary practice is the same as what is in dispute under the substantive law.

B. Qualifications

All experts must be "qualified."¹¹⁰ Rule 702, however, defines qualifications broadly. It does not limit experts to those with advanced degrees, but rather states that a witness can be qualified as an expert on the basis of "knowledge, skill, experience, training, or education."¹¹¹ Which type or types of qualifications are necessary for particular experts obviously depends on the subject matter of the proffered testimony.¹¹² But it also may depend upon whether the testimony is framework or diagnostic evidence.

1. Framework testimony.

When determining the qualifications of a framework expert under Rule 702, the rule's categories of knowledge, training, and

¹⁰⁸ See Thomas B. Newman and Michael A. Kohn, *Evidence-Based Diagnosis* 3 (Cambridge 2009) ("[T]he reason for diagnosis is to make treatment decisions[;] . . . diagnosing disease is important for treatment decisions because there are treatments that are beneficial in those who have a disease and not beneficial in those who do not.").

¹⁰⁹ *Id.*

¹¹⁰ FRE 702.

¹¹¹ FRE 702.

¹¹² See *Lujano v Town of Cicero*, 2011 WL 6097719, *3 (ND Ill) ("Ultimately, 'whether a witness is qualified as an expert can only be determined by comparing the area in which the witness has superior knowledge, skill, experience, or education with the subject matter of the witness's testimony.'"), quoting *Carroll v Otis Elevator Co*, 896 F2d 210, 212 (7th Cir 1990).

education are likely to play a disproportionate role. We refer to these three categories generally under the rubric of “knowledge,” since the ultimate purpose of education and training is to impart knowledge. On the other hand, the other two bases for expertise noted in Rule 702—skill and experience—are never, by themselves, sufficient to support testimony about an empirical framework.

General knowledge is crucial for the framework expert. Empirical frameworks inevitably transcend one person or one group of researchers.¹¹³ Because science is a cumulative enterprise, often involving many different research teams and, ideally, many different research designs, framework experts learn of the work of their fields primarily through education and training, which will usually be reflected in the academic degrees the expert possesses. In most areas of science, an expert’s background training also enables him or her to pursue a lifetime of further education. Thus courts should expect experts to have an abundant and up-to-date knowledge of the research in their fields.

The relevance of skill and experience in informing a scientific expert’s knowledge of empirical frameworks is much less clear. Of course, if the expert has conducted the research him or herself, skill in research methodology is an essential qualification. But many framework experts report the data rather than create them,¹¹⁴ in which case they need only have the knowledge necessary to critique the research methodology of the studies relied upon.

The role of experience is also ambiguous in the framework setting. The fundamental value of scientific research that informs framework testimony comes from the fact that it is not based on one person’s subjective experience or judgment. Thus, modern science puts little weight on the value of inductive experience for reaching generally applicable statements about some phenomenon.

But experience can still be indirectly useful to the researcher. For example, consider the situation of a psychiatrist interested in studying the alleged phenomenon of repressed memories.

¹¹³ See *Association for Molecular Pathology v United States Patent and Trademark Office*, 669 F Supp 2d 365, 377–78 (SDNY 2009) (describing the use of multiple research teams to study the possibility of the correlation between genetics and an increased risk of cancer).

¹¹⁴ See, for example, *Tillman v State*, 354 SW3d 425, 438 n 10 (Tex Crim App 2011) (admitting expert testimony describing eyewitness testimony even though the expert did not conduct the studies himself).

As a practicing psychiatrist, the expert might have come into contact with patients who claim that they have “discovered” memories that were once “repressed.”¹¹⁵ Without a methodologically sound research agenda, no amount of experience with individual patients is sufficient to demonstrate the empirical framework issue of whether the theory of repressed memories is valid. The psychiatrist’s patients are merely case studies and do not permit him or her to say whether the underlying phenomenon truly exists. But the exposure to patients still has value to a researcher. Together with his or her education and training, it will allow the expert to better formulate research designs that can test the posited theory.

2. Diagnostic testimony.

Like framework experts, diagnostic experts must be qualified on the basis of their “knowledge, skill, experience, training, or education.”¹¹⁶ But in the diagnostic setting much more emphasis is placed on skill and experience than on knowledge about the nuances of the scientific research. The diagnostic expert’s most important qualification concerns his or her ability to filter framework evidence through his or her skill set and experience for the benefit of the jury.

Every diagnostic expert must possess the relevant up-to-date education and training in the scientific area in which they will testify. But, in contrast to framework experts, who will focus solely on the relevant data, diagnostic experts should also have comprehensive training in forensic issues, because they are more likely to be asked questions that zero in on legal matters.¹¹⁷ For example, mental-health professionals who seek to testify about mental state in a criminal or civil case should not only have certain degrees (an MD in the case of a psychiatrist; typically a

¹¹⁵ See *Clark v Edison*, 881 F Supp 2d 192, 215 (D Mass 2012) (permitting such testimony).

¹¹⁶ FRE 702.

¹¹⁷ Under Federal Rule of Evidence 704 experts may give testimony that “embraces an ultimate issue to be decided by the trier of fact.” *Wade v Haynes*, 663 F2d 778, 783 (8th Cir 1981), *aff’d* 461 US 30 (1983). Our own view is that ultimate issue testimony ought to be avoided unless it is clearly based on the “specialized knowledge” required by Rule 702. See Christopher Slobogin, *The “Ultimate Issue” Issue*, 7 *Behav Sci & L* 259, 262 (1989). Be that as it may, diagnostic experts are routinely asked to give this type of testimony in some contexts. See, for example, Patricia A. Zapf, et al, *Have the Courts Abdicated Their Responsibility for Determination of Competency to Stand Trial to Clinicians?*, 4 *J Forensic Psychology Prac* 27, 42 (2004) (finding that virtually all judges prefer or demand ultimate issue testimony on competency).

PhD in clinical psychology for a psychologist), but also know the applicable legal tests (that is, the jurisdiction's insanity formulation or commitment criteria), knowledge which is best gained through participation in a joint-degree program in law, a forensic residency, or the process of obtaining board certification in forensic mental-health law.¹¹⁸

Moreover, diagnostic experts are much more likely than framework experts to need training not only with respect to a particular body of knowledge, but also with respect to particular skills. A primary benefit of diagnostic experts is their ability to gather information that nonexperts, or even framework experts, would find difficult to obtain. For instance, a medical doctor knows how to conduct tests that can reveal physical symptoms. A firearms expert is proficient at measuring striations on a bullet or gun barrel. But without training in the requisite skills, these individuals cannot claim to be diagnostic experts.

Experience is also important for diagnostic expertise, again more so than with framework expertise. More specifically, experience in the purported area of expertise is crucial. Focusing on the mental-health-professional example, neither education about the relationship of mental illness and crime nor general experience with evaluating and treating people with mental illness should be sufficient if the expert is proffered to testify about insanity. Rather, in addition to the relevant training, experience with evaluating and treating people who are mentally ill *and* charged with criminal offenses should be required.¹¹⁹

C. Scientific Validity

If *Daubert* wrought a revolution in expert evidence, a key element of the upheaval was the Court's mandate that scientific expert testimony "must be 'scientific . . . knowledge.'"¹²⁰ This "requirement," the Court explained, "establishes a standard of evidentiary reliability . . . based upon *scientific validity*."¹²¹ Thus, in tasking trial court judges to be gatekeepers, the Court mandated

¹¹⁸ See Gary B. Melton, et al, *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers* 24 (Guilford 3d ed 2007) ("Mental health professionals should not perform evaluations of competency to stand trial without knowledge of the standard.")

¹¹⁹ See *id* at 23–24 (stating that clinicians should have "experience in the relevant area" and giving examples).

¹²⁰ *Daubert*, 509 US at 589–90, quoting FRE 702.

¹²¹ *Daubert*, 509 US at 590 & n 9.

evaluation of the methods and principles underlying proffered expert testimony to assure its scientific validity. In order to assist judges in completing this daunting task, the Court in *Daubert* set forth the factors noted earlier: (1) whether the theory or technique can be, and has been, tested, (2) whether there is an acceptable error rate for the evidence and whether there are adequate standards for controlling the technique's operation, (3) whether the theory or technique has been subjected to peer review and publication, and (4) whether there is "widespread acceptance" of the theory or technique.¹²² *Kumho Tire* added a fifth factor: (5) whether the expert "employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."¹²³ As *Daubert* itself cautioned, these factors should not be treated as a recipe that can provide a definitive solution to every question of scientific validity.¹²⁴ But they do anticipate most of the inquiries that arise in connection with the admissibility of expert testimony.

Two preliminary points should be made before examining how these factors apply to framework and diagnostic testimony. First, assessing the validity of empirical framework evidence will generally be considerably easier than evaluating the validity of particularized testimony intended to be diagnostic. Most scientific evaluations of validity are of group effects, not individual results. Second, as we emphasized earlier, the validity of the relevant empirical framework is a prerequisite for the admissibility of diagnostic testimony. A particular case that is the subject of the adjudication cannot be said to be an instance of a general framework if the science supporting that framework does not exist in the first place.

1. Testability.

Testability is the *sine qua non* of basic science. On this point, the *Daubert* Court cited the philosopher of science Karl Popper, who asserted that testability distinguishes scientific statements from pseudoscientific or nonscientific statements.¹²⁵

¹²² *Id.* at 592–94.

¹²³ *Kumho Tire*, 526 US at 152.

¹²⁴ *Daubert*, 509 US at 594 ("The inquiry envisioned by Rule 702 is, we emphasize, a flexible one.").

¹²⁵ The criterion of falsifiability provides that "[a] statement or theory is . . . falsifiable if and only if there exists at least one potential falsifier—at least one possible basic

As the Court noted, Popper believed that “[t]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.”¹²⁶ The hallmark of scientific statements is that they are vulnerable to refutation. They also permit what Popper called intersubjective testability, which modern scientists might refer to simply as replication.¹²⁷

a) *Testability and framework testimony.* *Daubert* stated that the research hypotheses upon which expert testimony is based must not only be testable, but must also have been adequately tested.¹²⁸ Testing in science comes in myriad forms, ranging from highly controlled laboratory experiments to qualitative interviews. Very often, disparate methodologies are employed to test the same basic hypothesis. For example, the hypothesis that silicone implants can cause autoimmune illnesses was tested using multiple methods, including at the cellular level, employing laboratory animals, and by epidemiological methods on human populations.¹²⁹ Indeed, a key consideration in assessing the adequacy of the testing is whether different research methods were used and whether they converged on the same answer.¹³⁰

Thus, courts judging validity in a given case will often find that they need to assess numerous studies of various types. Consider, for example, the question whether violent video games can cause children to be violent. The hypothesis that there is a causal connection between violent media and increased violence of children has been the subject of considerable research attention using widely varying methods, including observational studies,

statement that conflicts with it logically.” Karl Popper, *Realism and the Aim of Science* xx (Rowman and Littlefield 1983) (W.W. Bartley III, ed).

¹²⁶ *Daubert*, 509 US at 593, quoting Popper, *Conjectures and Refutations* at 37 (cited in note 35).

¹²⁷ Karl Popper, *The Logic of Scientific Discovery* 44 (Basic Books 1961). See also Jason M. Chin, *Psychological Science’s Replicability Crisis and What It Means for Science in the Courtroom*, Psychology, Pub Pol & L (forthcoming 2014), online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412985 (visited May 21, 2014).

¹²⁸ *Daubert*, 509 US at 593 (“[A] key question to be answered in determining whether a theory or technique is scientific knowledge [is] . . . whether it can be (and has been) tested.”) (emphasis added). See also, for example, *Miller v Pfizer, Inc*, 196 F Supp 2d 1062, 1072 (D Kan 2002) (dismissing an expert’s hypothesis—that the drug Zoloft induces suicidal ideation—because it was testable but had not been adequately tested).

¹²⁹ See generally Jack W. Snyder, *Silicone Breast Implants*, 18 J Legal Med 133 (1997) (describing these and other types of studies).

¹³⁰ This principle is referred to as “convergent validity.” See William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* 350 (Houghton Mifflin 2d ed 2001).

behavioral experiments, and brain imaging.¹³¹ Courts confronting an expert opinion that violent video games can cause increased violence should have to consider the weight of all of this evidence before making an admissibility decision.¹³²

One conceptualization of the weight-of-the-evidence approach in the context of determining causation at the level of empirical frameworks comes from Austin Bradford Hill.¹³³ Hill described a set of nine factors that can assist scientists (and thus courts) in making judgments regarding general causation.¹³⁴ The Hill guidelines, and more generally the use of a “weight of the evidence” methodology, have gained some traction among courts.¹³⁵

This is a good trend. However, weight-of-the-evidence methods are also elastic and prone to strategic misuse. As Professors David Kaye, David Bernstein, and Jennifer Mnookin have explained, this method requires courts to examine “[t]he nature of the studies in each case, the plausibility of the extrapolations from them, and the known soundness of the basic theory—in sum, the expert’s causal reasoning”¹³⁶ Unfortunately, courts sometimes use the weight-of-the-evidence phraseology with no substantive evaluation of the evidence, much less any explanation

¹³¹ See generally Christopher J. Ferguson, *Violent Video Games and the Supreme Court*, 68 *Am Psychologist* 57 (2013).

¹³² This lesson was on display in *Brown v Entertainment Merchants Association*, 131 S Ct 2729, 2739 (2011) (dismissing studies purporting to indicate a connection between exposure to violent video games and harmful effects on children because they merely “show at best some correlation between exposure to violent entertainment and minuscule real-world effects, such as children’s feeling more aggressive or making louder noises in the few minutes after playing a violent game than after playing a nonviolent game”).

¹³³ Austin Bradford Hill, *The Environment and Disease: Association or Causation*, 58 *Proceedings Royal Socy Med* 295 (1965). See also Thomas A. Glass, et al, *Causal Inference in Public Health*, 34 *Ann Rev Pub Health* 61, 63 (2013).

¹³⁴ The nine factors he identified are (1) temporal relationships (in other words, if A is thought to cause B, A must precede B); (2) strength (that is, the stronger the association the more likely it is cause and effect); (3) dose-response (that is, causation gains support if the higher the dose the greater the effect); (4) consistency (that is, results are consistent across methods); (5) plausibility (that is, the mechanism believed to underlie causation is plausible); (6) consideration of alternative explanations (that is, other causes for the effect have been ruled out); (7) experiment (that is, causation judgments are higher if the research used experimental designs); (8) specificity (that is, the effect is specific to the putative cause); and (9) coherence (that is, the relationship is consistent with other research findings in the field). Hill, 58 *Proceedings Royal Socy Med* at 295–99 (cited in note 133).

¹³⁵ See *Dunn v Sandoz Pharmaceuticals Corp*, 275 F Supp 2d 672, 678–79 (MD NC 2003).

¹³⁶ David H. Kaye, David E. Bernstein, and Jennifer L. Mnookin, *The New Wigmore: A Treatise on Evidence: Expert Evidence* § 10.5.1 at 98 (Aspen Supp 2013).

of its weight.¹³⁷ Experts should be required to do more than simply assert the mantra “weight of the evidence.” They must specifically explain the component parts of the research evidence advanced to support their testimony and specify the weight they have accorded to each part.

b) Testability and diagnostic testimony. Testing the basis of diagnostic testimony is often a difficult enterprise, because by definition such testimony is always specific to a particular case. While a diagnostic opinion can certainly be informed by research and the confidence level associated with it can sometimes even be quantified, ultimately whether it is reliable in the legal sense can be tested only through some sort of feedback loop that indicates whether the expert was right or wrong. As Professors Daniel Kahneman and Gary Klein advise, intuitive expertise can be developed only in environments that have “high validity” (that is, useable outcome measures) and provide “an adequate opportunity to practice the skill” through repetitive trials.¹³⁸

Some diagnostic experts enjoy readily accessible feedback loops. For instance, diagnostic experts in various forensic fields—including handwriting, fingerprinting, and ballistics—can be subjected to proficiency testing, which measures their ability to identify correctly the relevant forensic product under a number of different scenarios.¹³⁹ The value of this testing, of course, depends on how closely it approximates the task at hand that is disputed in the courtroom. Similarly, the reliability and validity of various diagnostic instruments—ranging from breathalyzers and radar guns used by police to psychological tests and brain scans used by clinicians—must be established through testing and be recalibrated or renormed on a periodic basis.¹⁴⁰ In evaluating the validity of testimony from experts,

¹³⁷ See, for example, *Milward v Acuity Specialty Products Group, Inc.*, 639 F3d 11, 17–26 (1st Cir 2011), cert denied, 132 S Ct 1002 (2012) (reversing the district court exclusion of general causation expert testimony as an abuse of discretion, because experts claimed to use “weight of the evidence” methodology). See also David L. Faigman, et al, 3 *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 29:6 at 2–3 (West 2013) (criticizing *Milward* for confusing inference with a scientific methodology).

¹³⁸ Daniel Kahneman and Gary Klein, *Conditions for Intuitive Expertise: A Failure to Disagree*, 64 *Am Psychologist* 515, 520 (2009).

¹³⁹ For examples of such testing in the forensic arena, see Joseph L. Peterson, et al, *Crime Laboratory Proficiency Testing Research Program* 188–89 (National Institute of Law Enforcement and Criminal Justice 1978). See also *Florida v Harris*, 133 S Ct 1050, 1058–59 (2013) (discussing methods of conducting proficiency testing on drug-sniffing dogs).

¹⁴⁰ See, for example, John H. Blume, Sheri Lynn Johnson, and Christopher Seeds, *Of Atkins and Men: Deviating from Clinical Definitions of Mental Retardation in Death*

gatekeeping courts should assess the expert's success rate on, or the methodology of, these tests, when they exist.

Many diagnostic experts, however, cannot routinely test their conclusions in a given case in a scientific manner. For instance, consider an auto mechanic who testifies, based on his training and experience with Fords, that a fire in the Ford driven by the plaintiffs was probably the result of an oil leak onto the exhaust manifold when the driver braked hard. Assuming the fire destroyed the car, the mechanic cannot verify the accuracy of his testimony, and replicating the conditions of the fire is not possible. Or consider a psychiatrist who testifies in an insanity trial that, based on his experience evaluating people with serious mental illness, the defendant was substantially unable to appreciate the wrongfulness of his actions. Unfortunately, there is no gold standard for insanity; insanity, and whether a person's appreciation of wrongfulness was "substantial" in a particular case, is simply what the jury says it is, and thus the expert cannot know, in any absolute sense, whether the opinion is "correct."¹⁴¹ Next assume that a diagnostic expert opines that a particular convicted offender is a high risk for committing another crime in the near future. If the court agrees, the individual will be incarcerated or treated or both. Once again, the feedback loop is compromised. The outcome variable here (reoffending) is neither amorphous (as with insanity) nor difficult to ascertain (as it is with respect to the cause of car fires). But because of the intervention by the court, the law prevents a falsification procedure.

These examples illustrate the difficulty of measuring in a scientific way the methodology or opinions of diagnostic expertise in many types of cases. The necessary feedback loops simply do not exist. However, a diagnostic expert can at least attend to the *process* of accumulating and analyzing the relevant information. As developed further below, whether the expert follows the approved process can be a form of testing.

Penalty Cases, 18 Cornell J L & Pub Pol 689, 700–01 (2009) (discussing the renorming of IQ tests for determining whether a person is intellectually disabled for death-penalty purposes).

¹⁴¹ See generally Douglas Mossman, Dustin B. Wygant, and Roger O. Gervais, *Estimating the Accuracy of Neurocognitive Effort Measures in the Absence of a "Gold Standard"*, 24 Psychological Assessment 815 (2012). As one court stated, "Legal tests of criminal insanity are not and cannot be the result of scientific analysis or objective judgment. There is no objective standard by which such a judgment of an admittedly abnormal offender can be measured." *Sauer v United States*, 241 F2d 640, 648 n 27 (9th Cir 1957), quoting *Holloway v United States*, 148 F2d 665, 666 (DC Cir 1945).

2. Error rates and adequate standards.

In *Daubert*, the Court stated that “in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error . . . and the existence and maintenance of standards controlling the technique’s operation.”¹⁴² The concept of error—or unexplained variance—is central to both the general scientific research that informs framework testimony and the particular scientific techniques that might be employed to offer diagnostic testimony. Indeed, in science, error—conceptualized as what is yet unknown—fuels the scientific enterprise itself. While analysis of the “standards” maintained by the relevant discipline is a distinct issue, it is closely connected to the error-rate factor and will also be discussed here.

a) *Error rates and standards in framework testimony.* Inherent in the testability factor is an error rate. At the general level of science, the notion of error is not strictly reducible to a single “rate” or number. Yet error is an abiding consideration in scientific research.¹⁴³ Assume, for instance, an attempt to ascertain the toxic effects of Substance X. For ethical reasons a prospective randomized study of X’s toxic effects can be tested only on animals, which introduces the type of error we earlier discussed in connection with empirical fit. An alternative research method, epidemiological study of X’s effects on humans who have previously been exposed to it, introduces error because it occurs without random assignment, and could also encounter problems associated with inadequate sample size or inadequate comparison groups (that is, groups that have not been exposed to X).¹⁴⁴ Also, errors in research can have random or systematic effects. Random error would occur, for example, if researchers made mistakes in coding their data, such as recording that subjects given an experimental drug were given a placebo, and vice versa.¹⁴⁵ In contrast, systematic errors produce bias in a single

¹⁴² *Daubert*, 509 US at 594 (citations omitted). Although the Court mentioned error rates only briefly, this factor has become integral to the *Daubert* analysis and a considerable legal literature has been devoted to it. See generally D. Michael Risinger, *Whose Fault?—Daubert, the NAS Report, and the Notion of Error in Forensic Science*, 38 *Fordham Urban L J* 519 (2010).

¹⁴³ See Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs* at 401–04 (cited in note 130).

¹⁴⁴ *Id.* at 13–18.

¹⁴⁵ *Id.* at 401–02.

direction.¹⁴⁶ For example, if researchers rely on self-reports of exposure to X, and if some of the people asked are ill, “recall bias” about whether and when such exposure took place can arise.¹⁴⁷

As its linkage in *Daubert* suggests, the extent to which the relevant discipline maintains standards for evaluating the validity of its research techniques is closely associated with error rates. For example, most scientists consider a finding of differences in values (effect size) to be statistically significant if the result is demonstrably likely to occur by chance only one in twenty times, assuming no differences exist (that is, the null hypothesis is true).¹⁴⁸ The significance of statistically significant results can also be analyzed statistically, through conventions that determine whether a given effect size is small, medium, or large.¹⁴⁹ Using *Daubert’s* language, the significance of error rates can be evaluated in standard ways. Consider for example, a study that shows that, on a 5-point scale measuring ability to make appropriate prosocial judgments, juveniles who are 16 years of age score 2.77 on average while adults score on average 3.1, a result that, if valid, could be important in deciding the responsibility of juvenile offenders.¹⁵⁰ For this difference to be valid, however, statistical analysis must first show that the likelihood of obtaining this difference was 5 percent or less. Further, even assuming the difference is statistically significant, one might still want to know whether the effect size is considered small, medium, or large as a statistical matter, to provide some measure of whether the statistically significant results are of any *practical* significance.

Courts must be attuned to these methods of evaluating the worth of framework evidence, even if it means wading into unfamiliar statistical waters. Without error-rate information

¹⁴⁶ Id at 410–13.

¹⁴⁷ Karen Raphael, *Recall Bias: A Proposal for Assessment and Control*, 16 Intl J Epidemiology 167, 167 (1987).

¹⁴⁸ Unfortunately, some experts, and therefore courts, erroneously interpret this finding to mean that the “error rate” is 5 percent. See David Faust, et al, *The Admissibility of Behavioral Science Evidence in the Courtroom: The Translation of Legal to Scientific Concepts and Back*, 6 Ann Rev Clinical Psychology 49, 73 (2010).

¹⁴⁹ See Jacob Cohen, *A Power Primer*, 112 Psychology Bull 155, 157 (1992) (suggesting benchmarks for interpreting Cohen’s *d* effect sizes of .20 (designated a small effect size), .50 (medium), and .80 (large)).

¹⁵⁰ See Elisabeth Cauffman and Laurence Steinberg, *(Im)maturity of Judgment in Adolescence: Why Adolescents May Be Less Culpable than Adults*, 18 Behav Sci & L 741, 752 (2000) (reporting these results).

Daubert's command that courts evaluate the reliability of expert evidence cannot be accomplished, at least when an empirical framework is involved.

b) *Error rates and standards in diagnostic testimony.* Error-rate information is also sometimes available in connection with diagnostic testimony, especially when it is predicated on some type of instrument. For example, the Classification of Violence Risk (COVR), a software program that provides an actuarial violence-risk assessment that can be used to inform diagnostic testimony in civil-commitment cases, yields a prediction of violent behavior, as well as a statement of the confidence with which this prediction can be taken. An illustrative output from this instrument would be: Based on the data used to construct the Classification of Violence Risk, one can say with 95 percent confidence that between 20 and 32 percent of persons with the same score as [person's name] can be expected to commit a violent act toward another person in the next several months, with a best estimate of 26 percent.¹⁵¹

With other types of diagnostic testimony, error rates cannot be generated through the types of large-scale studies that resulted in the COVR. That does not mean that error rates cannot be determined in such situations, however. For instance, in *Kumho Tire*, the Court stated that, “[i]n certain cases, it will be appropriate for the trial judge to ask, for example, how often an engineering expert’s experience-based methodology has produced erroneous results”¹⁵² With this language, the Court recognized that error rates can sometimes be obtained even in cases in which diagnostic testimony is based on the expert’s purported knowledge alone; the expert can, for instance, simply keep track of his or her own success rate in the types of cases at issue.

However, as noted in our discussion of the testability factor, diagnostic expertise is not always amenable to even this type of verification process. Thus, diagnostic experts will not always be able to provide anything resembling error-rate information. In

¹⁵¹ See John Monahan, et al, *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence* 126–27 (Oxford 2001). See also Nicholas Scurich, John Monahan, and Richard S. John, *Innumeracy and Unpacking: Bridging the Nomothetic/Idiographic Divide in Violence Risk Assessment*, 36 *L & Human Behav* 548, 552 (2012) (reporting research on violence-risk assessment “asking empirically under what circumstances people apply group-level (i.e., actuarial) estimates to an individual member of the group”).

¹⁵² *Kumho Tire*, 526 US at 151.

such cases, the *process* by which experts arrive at their conclusions—what *Daubert* called “standards [for] controlling the technique’s operation” and what we will call “protocols”¹⁵³—can be subject to study, and, as a result, a form of error rate can be generated.

For instance, violence-risk-assessment guides tell clinicians the factors that studies have found are most strongly related to violence; if a diagnostic expert fails to consider those risk factors (such as age, gender, and prior violence), error can be said to have occurred. Similarly, in cases involving medical issues, professionals can identify best practices for conducting evaluations, and an expert’s failure to follow the approved structured format could be viewed as a type of error, with the number of departures constituting the error rate.¹⁵⁴ This type of case-specific error-rate analysis would be particularly relevant for a court attempting to ascertain whether the expert has followed an adequate evaluation procedure and is thus qualified to testify in the particular case.

How are these procedures—or, in diagnostic terminology, protocols—to be developed? In some instances of diagnostic expertise, researchers can assist in this endeavor. Protocols that can be subjected to scientific testing, such as, for example, polymerase chain reaction (PCR) methods that generate DNA profiles, can be tested and retested by cross-validating them on new sample populations. That process can fine-tune the appropriate standards to follow.

Protocols that cannot be easily subjected to scientific testing require a different approach to standard-setting. One obvious method is to place the burden on the relevant professional organizations to devise best-practice evaluation procedures. The Social Security Administration has developed a highly structured method of assessing disability for the purpose of determining benefits eligibility.¹⁵⁵ The medical profession has established protocols for conducting various types of tests, ranging from mammograms to brain scans,¹⁵⁶ and in mental health there are

¹⁵³ *Daubert*, 509 US at 594, citing *United States v Williams*, 583 F2d 1194, 1198 (2d Cir 1978).

¹⁵⁴ See Robert G. Bone, *Securing the Normative Foundations in Litigation Reform*, 86 BU L Rev 1155, 1160 (2006) (“Procedure reduces . . . the number of errors and thus the total harm that the system’s errors create.”).

¹⁵⁵ See generally 20 CFR § 404.1505 et seq.

¹⁵⁶ See Arnold J. Rosoff, *Evidence-Based Medicine and the Law: The Courts Confront Clinical Practice Guidelines*, 26 J Health Polit Pol & L 327, 343 (2001).

well-accepted protocols for conducting child-custody and insanity evaluations and structured instruments for evaluating the risk of an offender.¹⁵⁷ These types of best-practice, or evidence-based, standards should be highly relevant to the admissibility decision in diagnostic-evidence cases.¹⁵⁸

3. Peer review and publication.

The expectation that research has been published in a peer-reviewed journal is one shared by most mainstream scientists. But many fields on which courts rely—including the legal academy—do not share this view, and it must also be recognized that peer review can occur outside the publication process. Various versions of the peer review dynamic can be seen both in general fields of inquiry that inform framework testimony and in the particularized judgments that might be introduced in court as diagnostic testimony.

a) Peer review and publication in framework testimony. The basis for framework evidence should usually be accessible in a peer-reviewed journal.¹⁵⁹ But not all peer-reviewed journals are equal in status. Scientists judge the quality of scholarship, in part, by the selectivity of the journal in which it appears. This factor should be important for judges to consider too.¹⁶⁰ Yet, at the same time, peer review is no panacea, and many articles that survive it turn out to be invalid in time. Further, conventional scientific journals likely tend toward publishing conventional scientific research,

¹⁵⁷ See, for example, Kirk Heilbrun, Thomas Grisso, and Alan M. Goldstein, *Foundations of Forensic Mental Health Assessment* 5–10 (Oxford 2009); Thomas Grisso, *Forensic Evaluation of Juveniles* 153 (Professional Resource 1998); Jennifer L. Skeem and John Monahan, *Current Directions in Violence Risk Assessment*, 20 *Curr Dir in Psychological Sci* 38, 39 (2011).

¹⁵⁸ See Maxwell J. Mehlman, *Professional Power and the Standard of Care in Medicine*, 44 *Ariz St L J* 1165, 1216–24 (2012) (reporting the extent to which medical evidence-based guidelines have influenced court decisions and describing the difficulties with ensuring the guidelines are based on good, up-to-date data). See also Christopher Slobogin, *Proving the Unprovable: The Role of Law, Science, and Speculation in Adjudicating Culpability and Dangerousness* 64–71 (Oxford 2007) (describing and analyzing a proposal by Professor Daniel Fishman for establishing a computer-based method of accumulating psychological cases studies that can be used to hone procedures and establish professional standards).

¹⁵⁹ The prestigious journal *Nature* has an excellent description of the peer review process on its website. See *Peer-Review Policy* (Nature Publishing Group 2013), online at http://www.nature.com/authors/policies/peer_review.html (visited May 21, 2014).

¹⁶⁰ See, for example, *Daubert v Merrell Dow Pharmaceuticals, Inc.*, 43 F3d 1311, 1318 (9th Cir 1995) (noting the absence of publications by the plaintiffs' expert in scientific journals).

thus limiting the volume of revolutionary findings to be found in them. These considerations, on balance, support the *Daubert* Court's conclusion that peer review and publication is an important factor in assessing admissibility, but not a prerequisite to it.¹⁶¹

In considering the quality of the journals in which empirical framework information appears, judges would be well-advised to return to the first two factors the *Daubert* Court identified, testing and error rate. Publication in journals that do not share the values reflected in these first two factors should alert judges that the research must be scrutinized with particular care. To be sure, just as the Court did not establish publication as a prerequisite to admissibility, so too publication in nonrigorous journals—or, even, in student- (rather than peer-) edited law reviews¹⁶²—should not mean per se exclusion. But in both cases judges should use extreme caution before admitting testimony relying on them.

The principle of peer review can also be understood much more broadly, and more fundamentally, than merely getting a research study published. In *Daubert*, the Court observed that “submission to the scrutiny of the scientific community is a component of ‘good science,’ . . . because it increases the likelihood that substantive flaws in methodology will be detected.”¹⁶³ In practice, this aspect of peer review—detecting serious flaws in the design of studies—is not restricted to the evaluation of journal reviewers. It is the culture of science and thus also takes place before research is conducted (such as when a funding agency evaluates research proposals) and continues well past publication. Indeed, most of the “scrutiny of the scientific community”¹⁶⁴ takes place after publication. That is why *Daubert*'s discussion of peer review and publication takes pains to alert courts to the fact that publication in a peer-reviewed journal is no assurance of the soundness of a finding, and the lack of publication is no assurance that a study and its findings are not sound.¹⁶⁵

¹⁶¹ *Daubert*, 509 US at 593–94 (recognizing that publication “does not necessarily correlate with reliability,” but is nonetheless a “relevant . . . consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised.”).

¹⁶² See, for example, *Cicero v Borg-Warner Automotive, Inc.*, 163 F Supp 2d 743, 747–48 (ED Mich 2001) (noting that proffered expert, a law professor, had published only in student-edited law reviews). See also Shari Seidman Diamond and Pam Mueller, *Empirical Scholarship in Law Reviews*, 6 Ann Rev L & Soc Sci 581, 592–93 (2010).

¹⁶³ *Daubert*, 509 US at 593 (citations omitted).

¹⁶⁴ *Id.*

¹⁶⁵ See *id.* at 594.

b) *Peer review and publication in diagnostic testimony.* While diagnostic expertise is particularized to the case and thus is not normally the subject of a published report (at least prior to adjudication), it can nonetheless be subject to a version of peer review. Indeed, the classic form of peer review is the second opinion. Ensuring that a colleague agrees with one's diagnostic assessment, or at least the method of arriving at it, can provide some assurance of validity.¹⁶⁶

The use of second opinions as a peer review mechanism must be approached with caution, however. Second opinions are sometimes overeagerly touted as a form of peer review in various forensic-identification fields, including, among others, latent-fingerprint identification. Indeed, the basic methodology in latent-print analysis contains four stages, one of which (Verification) is explicitly a double-check stage that follows Analysis, Comparison, and Evaluation of an unknown print (thus the entire process is given the acronym ACE-V).¹⁶⁷ Critics of the ACE-V method question whether verification operates in practice as well as it is idealized in theory.¹⁶⁸ For instance, verifiers in the forensic sciences typically consider only the one comparison they are given (in other words, the conclusion reached by the original analyst), they know the identity of that analyst, and they have background case information regarding the comparison in question.¹⁶⁹ Thus, like peer review and publication, courts should ensure that diagnostic second opinions (or verifications) are effective. They should confirm that these reviews are not mere rubber stamps of the original finding and are not overly influenced by knowledge of the identity of the preceding analysts or biased by extraneous background information.

Peer review of a diagnostic opinion via publication is also possible, at least *ex post*, in a way that can aid future diagnostic experts. For instance, case studies can be described in journals that publish only articles vetted by other professionals. Ideally,

¹⁶⁶ See *Rey v Texas*, 897 SW2d 333, 338 (Tex Crim App 1995) (holding that pathology, in general, and psychiatry in particular, are both subspecialties of medicine that are not exact science like mathematics and therefore, there is a need for a second opinion).

¹⁶⁷ See Jennifer L. Mnookin, et al, *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L Rev 725, 734–35 (2011); National Research Council, *Strengthening Forensic Science* at 137–38 (cited in note 90).

¹⁶⁸ See Sandy L. Zabell, *Fingerprint Evidence*, 13 J L & Pol 143, 178 (2005) (“ACE-V is an acronym, not a methodology.”).

¹⁶⁹ See Glenn Langenburg, Christophe Champod, and Pat Wertheim, *Testing for Potential Contextual Bias Effects during the Verification Stage of the ACE-V Methodology When Conducting Fingerprint Comparisons*, 54 J Forensic Sci 571, 571 (2009).

these vetting professionals would be “experts on experts.” Several fields have such publications.¹⁷⁰ Ultimately, in areas in which feedback loops are hard to develop, determining who fits in this expert-on-experts category will depend to some extent on who appears to be the most qualified—that is, those who are the best educated and trained, who have the most experience, and who most closely follow approved evaluation procedures in their everyday practice.¹⁷¹ If and when courts use this method of evaluating the peer review criterion in the diagnostic setting, they should make some inquiry into this aspect of the publication process.

4. General acceptance of the method.

Both *Frye* and *Daubert* incorporate general acceptance of the method or theory undergirding the expert’s testimony in admissibility. *Frye* makes the issue dispositive, while *Daubert* relegates it to the last of four factors. General acceptance can be a useful consideration in determining admissibility of both framework and diagnostic testimony if, as *Daubert* apparently intended, courts look to experts in the relevant field rather than merely canvass the views of other courts.¹⁷²

a) *General acceptance of framework testimony.* Framework evidence, by definition, comes from a field of expertise, although perhaps a very new or small one. The key difficulty here involves identifying what field should be selected for review. Very often fields are defined by self-interested practitioners or established guilds. For example, if a court asks experts in the areas of polygraphs, bite marks, bullet lead, or hair identification about general acceptance, it would likely hear a chorus of consensus, though each of these areas of claimed expertise has been thoroughly discredited.¹⁷³ Similarly, for example, handwriting-identification experts are a closed society and, since their livelihoods depend on

¹⁷⁰ For instance, medical journals often include such case studies. See, for example, Sally Macintyre, et al, *Using Evidence to Inform Health Policy: Case Study*, 322 *British Med J* 222, 224 (2001).

¹⁷¹ See Paul S. Appelbaum, *Policing Expert Testimony: The Role of Professional Organizations*, 53 *Psychiatric Services* 389, 390 (2002).

¹⁷² See, for example, Saul M. Kassin, et al, *On the “General Acceptance” of Eyewitness Testimony Research: A New Survey of the Experts*, 56 *Am Psychologist* 405, 413–15 (2001).

¹⁷³ See National Research Council, *Strengthening Forensic Science* at 150–52, 155–61, 173–76 (cited in note 90). See also Michael J. Saks, *Forensic Identification: From a Faith-Based “Science” to a Scientific Science*, 201 *Forensic Sci Intl* 14, 15 (2010).

the acceptance of their expertise, few dissenting voices are likely to be heard.¹⁷⁴ In other situations, the opposite problem occurs: scientific evidence involves more than one academic or professional discipline. For instance, should the validity of the battered woman syndrome be assessed by asking psychiatrists, research psychologists, or clinical psychologists?

Measuring general acceptance, then, largely involves a process of deciding whose noses to count. In deciding this issue, courts should keep in mind two basic considerations. First, they must consult a broad enough spectrum of scientists to reach those who are not entirely invested in the expertise. “Investment” in this context could refer to financial interests, but is intended to be considerably broader, and include professional and ideological. Professional overinvestment might include a researcher who is a leading figure in the field but whose life’s work depends on acceptance of the expertise. Ideological investment might include a researcher whose judgment about the validity of an empirical framework will be influenced by its ability to further a desired policy outcome.¹⁷⁵ While financial, professional, or ideological investment should not necessarily disqualify an expert from addressing the general acceptance question, courts should strive to get input from those whose judgments are relatively free of such influence.¹⁷⁶

The second basic consideration for courts to consider in measuring general acceptance—one that is particularly important in framework-evidence cases—is in some tension with the first, because it requires that those surveyed truly understand the research basis for the empirical framework. Thus, for example, while the first consideration would require courts interested in the acceptance of polygraph technology to go beyond polygraph examiners, this second consideration would require courts to take note of the fact that other possibly knowledgeable professionals (for example, psychologists and statisticians)

¹⁷⁴ See National Research Council, *On the Theory and Practice of Voice Identification* 68–69 (National Academy of Sciences 1979) (indicating that a group composed of acoustical engineers, physiologists, statisticians, and others gave a less favorable assessment of the technique than the narrower range of developers of the technique).

¹⁷⁵ See Dan M. Kahan, Hank Jenkins-Smith, and Donald Braman, *Cultural Cognition of Scientific Consensus*, 14 *J Risk Rsrch* 147, 157–66 (2011); Philip E. Tetlock, Gregory Mitchell, and L. Jason Anastasopoulos, *Detecting and Punishing Unconscious Bias*, 42 *J Legal Stud* 83, 93–102 (2013).

¹⁷⁶ See, for example, Katrina Rufino, et al, *When Experts Disagreed, Who Was Correct? A Comparison of PCL-R Scores from Independent Raters and Opposing Forensic Experts*, 36 *L & Human Behav* 527, 535 (2012).

might not be familiar with it. In some cases independent groups of scientists, such as those sponsored by the National Academies of Science,¹⁷⁷ will have reviewed the validity of certain kinds of expert testimony. But in most areas of scientific evidence, courts will not have the benefit of such reports and will have to do their best to examine the views of the true believers in light of whatever critical commentary is available from respected scholars outside the field.

b) General acceptance of diagnostic testimony. For diagnostic experts whose expertise derives primarily from scientific research in their fields, evaluating the general acceptance of their testimony will be very similar to evaluating the general acceptance of the testimony of framework experts. Thus, the points made in the prior Section about the difficulty of delineating the appropriate field of reference apply here as well. As with framework evidence, the existence of independent bodies that have identified, based on scientific study, best practices in the field can be extremely useful to courts. For example, if the Institute of Medicine of the National Academy of Sciences issues a consensus report on how post-traumatic stress disorder (PTSD) should be clinically assessed among combat veterans of the wars in Iraq and Afghanistan,¹⁷⁸ this should go very far in determining whether the manner in which a diagnostic expert evaluated a criminal defendant claiming combat-induced PTSD followed a generally accepted protocol. Of course, the expert must also demonstrate that he or she employed the accepted protocol in the approved manner.

General acceptance is more difficult to evaluate when the diagnostic testimony is not explicitly research-based. In discussing the general acceptance factor in *Kumho Tire*, the Supreme Court stated that “it will at times be useful to ask even of a witness whose expertise is based purely on experience, say, a perfume tester able to distinguish among 140 odors at a sniff, whether his preparation is of a kind that others in the field would recognize as acceptable.”¹⁷⁹ Given the lack of quantified

¹⁷⁷ See, for example, National Research Council, *The Polygraph and Lie Detection* (National Academies 2003), online at http://www.nap.edu/catalog.php?record_id=10420 (visited May 21, 2014).

¹⁷⁸ As it has. See generally Institute of Medicine, *Treatment for Posttraumatic Stress Disorder in Military and Veteran Populations: Initial Assessment* (National Academies 2012). A second report on this topic is scheduled to be released by the Institute of Medicine (IOM) in 2014.

¹⁷⁹ *Kumho Tire*, 526 US at 151.

information about experience-based diagnostic expertise, with the concomitant difficulty of testing the resulting opinions, generating error rates, and constructing a useful peer review process, the inquiry into general acceptance may be the most important in considering the admissibility of this type of testimony.

As always, the court must not only assure that a practice is generally accepted but also that the practice fits the case at hand. For instance, perhaps the diagnostic expert diligently follows a generally accepted structured interview format for determining the presence of PTSD, but the precise legal issue in the case in which he or she is asked to testify is whether PTSD can cause a failure to appreciate the wrongfulness of one's actions. Unless the expert also resorted to a structured format for evaluating insanity,¹⁸⁰ a court might justifiably decide that the evaluation process is not one that is generally accepted for the case before it.

5. Same intellectual rigor.

In *Kumho Tire*, the Court reiterated "*Daubert's* gatekeeping requirement" and repeated the four factors just canvassed.¹⁸¹ But, as we've noted, it also stated that judicial gatekeeping would "make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."¹⁸² This standard has proved to be quite popular with lower courts¹⁸³ and is worth separate consideration.

a) *Same intellectual rigor in framework testimony.* Because, like the general acceptance test, it relies on practices in the relevant field, the same-level-of-intellectual-rigor test presents similar dangers to those that were presented by the *Frye* test. If the fields themselves are not valid, this deference leads courts to accept bad framework science. Closely related to the possibility that a field has little intellectual rigor is the not unusual situation in which the field does not exist outside the courtroom. Most of the forensic-identification sciences, for example, have only one

¹⁸⁰ For such a format, see Ira K. Packer, *Evaluation of Criminal Responsibility* 30–31 (Oxford 2009).

¹⁸¹ *Kumho Tire*, 526 US at 152.

¹⁸² *Id.*

¹⁸³ See, for example, *Jenkins v Bartlett*, 487 F3d 482, 489 (7th Cir 2007).

market, the law.¹⁸⁴ For these specialties, the expert can always testify that he or she uses the same level of intellectual rigor in the courtroom that characterizes the practice in the relevant field, because the two are the same.

An object lesson involving the tension between framework testimony offered in court and professional standards comes from the expert testimony in *Pippen v Iowa*,¹⁸⁵ a class action alleging that Iowa's failure to monitor its state hiring system improperly allowed "implicit bias" to affect personnel decisions to the detriment of African American job applicants.¹⁸⁶ Professor Anthony Greenwald, a well-known research psychologist, testified for the plaintiffs that "based [on] his nationwide research, roughly three-quarters of the people able to affect hiring decisions in Iowa's state government have an implicit bias that makes them favor whites over blacks."¹⁸⁷ When asked whether he could "prove" that bias led to the statistically low number of African Americans employed by the state, he stated, "I would be reluctant to make that contention in the journal publication context, because I think that requires a higher standard of proof."¹⁸⁸ Instead, Greenwald continued, "I can say that it is plausible that implicit bias is a cause of discrimination in the state of Iowa. I regard that as a plausible hypothesis that I would love to test."¹⁸⁹ Greenwald's assertion that "the journal publication context" requires "a higher standard of proof" than the courtroom starkly indicates a failure to apply in the courtroom "the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."¹⁹⁰

b) Same intellectual rigor in diagnostic testimony. In the diagnostic context, if the requirement that the expert employ "the same level of intellectual rigor that characterizes the practice of an expert in the relevant field" is to mean something different

¹⁸⁴ See *Daubert*, 43 F3d at 1317 n 5 ("Fingerprint analysis, voice recognition, DNA fingerprinting and a variety of other scientific endeavors closely tied to law enforcement may indeed have the courtroom as a principal theatre of operations.").

¹⁸⁵ No. LACL107038, slip op (Iowa D Ct Polk County, Apr 17, 2012).

¹⁸⁶ *Id.* at 52.

¹⁸⁷ Jeff Eckhoff, *Update: Polk Judge Rules That "Uncommon Approach" to Class-Action Racial Bias Case Fell Short of the Law*, Des Moines Register Blog (Des Moines Register Apr 17, 2012), online at <http://blogs.desmoinesregister.com/dmr/index.php/2012/04/17/polk-co-judge-rejects-class-action-lawsuit-alleging-that-state-government-discriminated-against-blacks> (visited May 21, 2014).

¹⁸⁸ *Id.*

¹⁸⁹ *Id.*

¹⁹⁰ *Kumho Tire*, 526 US at 152.

than is already captured by the general acceptance and peer review requirements, it should focus on the quality of the expert's analysis of whatever information the expert acquires through his or her (generally accepted) assessment process. In this regard, the "same intellectual rigor" factor is analogous to the concept of custom in tort law.¹⁹¹ Custom in the context of tort law means that the relevant party acted as other people—here, other professionals—would have acted in similar circumstances. "[C]omplying with custom confirms that the actor has behaved in the ordinary way."¹⁹² According to Professor Kenneth S. Abraham, "[a]dmitting custom evidence reflects the idea that recurring patterns of conduct have a bearing on what constitutes reasonable care."¹⁹³

As described earlier, it is customary in the ordinary practice of medicine and related fields (for example, clinical psychology) for professionals to make individual diagnostic judgments derived from group-based data. Likewise, it is *not* customary in the ordinary practice of sociology, epidemiology, anthropology, and related fields (for example, cognitive and social psychology) for professionals to make individual diagnostic judgments derived from group-based data.¹⁹⁴ In the law of evidence as in tort law, being informed of the individual inferences drawn by diagnostic experts such as physicians and clinical psychologists

has the potential to acquaint the jury with unfamiliar fields of activity, to place the conduct of the parties in its proper social or economic context, and to help the jury get its bearings. Custom evidence informs the jury about what happens in the real world, and thereby may enhance the accuracy and reliability of jury decisionmaking.¹⁹⁵

¹⁹¹ While we focus here on the use of custom in tort, custom is a central concept in many diverse areas of law. As the Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 13 (2010) states, "Custom plays a powerful role in the law generally. In international law, under certain circumstances custom can be the actual source of legal obligations. In contract law, customs are often understood to be implied terms in individual contracts." *Id.* at § 13, comment a.

¹⁹² *Id.* See also generally *The Duke Project on Custom and Law* (Duke University 2013), online at <http://law.duke.edu/customlaw> (visited May 21, 2014).

¹⁹³ Kenneth S. Abraham, *Custom, Noncustomary Practice, and Negligence*, 109 Colum L Rev 1784, 1785 (2009).

¹⁹⁴ We discuss the contrary view of Fiske and Borgida. See note 232 and accompanying text.

¹⁹⁵ Abraham, 109 Colum L Rev at 1787 (cited in note 193).

Two problems arise in connection with this inquiry into customary rigor, however. First, the legal query may be so unique that experts in the relevant field never have occasion to address it (as in our earlier example involving testimony about the etiology of acute coronary ischemia) or so empirically fraught that experts in the field do not believe it should be addressed (as in the eyewitness-testimony setting). The second problem arises when experts in the field do engage in the relevant analysis but their inferential reasoning is opaque and thus not easily subject to scrutiny. The Supreme Court's perfume tester might be an example of this phenomenon.¹⁹⁶ The question that arises here is whether diagnostic testimony that cannot be explained should be admissible. We earlier suggested it should be, if reliability can be gauged through rigorous proficiency testing or a similar procedure. Other commentators have taken a different view.¹⁹⁷

D. Added Value ("Helpfulness")

Even if expert testimony is material, the expert is qualified, and the basis of the testimony is valid (the three evidentiary factors discussed to this point), it is not admissible if it does not assist the fact finder in making the relevant decision. Rule 702 states that expert testimony must "help" the jury,¹⁹⁸ and *Daubert* reiterated that point.¹⁹⁹ But expert testimony does not have to transform the fact finder from "incompetence" to "competence" when it comes to understanding the evidence or determining a fact in issue. It simply has to add value, that is, make a positive contribution to the process of determining facts.²⁰⁰

At the same time, helpfulness is usefully distinguished from advocacy and argument. Expert testimony along the lines of

¹⁹⁶ See *Kumho Tire*, 526 US at 151.

¹⁹⁷ See, for example, Ronald Allen, *Expertise and the Supreme Court: What Is the Problem?*, 34 Seton Hall L Rev 1, 9 (2003) (stating that "my present view is that a person who cannot explain the basis of testimony in an accessible fashion or explain how it can be verified ought not be allowed to testify").

¹⁹⁸ FRE 702:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue.

¹⁹⁹ *Daubert*, 509 US at 591.

²⁰⁰ See Kaye, Bernstein, and Mnookin, *The New Wigmore* § 2.1.2 at 39 (cited in note 136) (stating that expert testimony "must allow the expert to provide useful information or perspective that the jurors do not already have").

“here’s how I would vote if I were on the jury” is not admissible.²⁰¹ Neither is expert testimony that merely mimics lawyers’ arguments.²⁰²

1. Framework testimony.

The primary added value of framework testimony is that it provides a context in which a particular case might be understood. Given our focus on G2i, this added value can be understood as serving two basic purposes. The first purpose is to give jurors a background with which they might better understand facts in the particular case at hand. The second purpose is to provide a subsequent expert witness with the general framework in which to situate his or her proffered diagnostic testimony.

When framework testimony is admitted without accompanying diagnostic expert testimony, its principal purpose is educative. It allows jurors to understand a facet of the world that they know little about, or about which they might hold certain misconceptions. At common law, this question concerned whether the subject matter was “beyond the ken” of the average layperson.²⁰³ Modern rules of evidence have a less demanding threshold for added value, but courts continue to insist that expert testimony provide assistance beyond what lay jurors could do on their own.²⁰⁴

Consider, for example, the phenomenon of false confessions, which has been the subject of considerable scientific research. Most jurors are unlikely to know about the subject and, indeed, the phenomenon itself seems incredible. Why would anyone falsely confess? Yet the DNA exoneration cases demonstrate that false confessions are not infrequent.²⁰⁵ Framework testimony on

²⁰¹ See FRE 704, Advisory Committee Notes to the 1972 Proposed Rules (“Under Rules 701 and 702, opinions must be helpful to the trier of fact, and Rule 403 provides for exclusion of evidence which wastes time. These provisions afford ample assurances against the admission of opinions which would merely tell the jury what result to reach . . .”).

²⁰² See Jack B. Weinstein and Margaret A. Berger, 4 *Weinstein’s Federal Evidence* 702-41 at § 702.03[2][a] (Lexis 2d ed 2013) (“Proffered expert testimony should be excluded when it will not help the trier of fact to any degree beyond the assistance that the lawyers representing the parties could provide during their closing arguments.”).

²⁰³ See Kaye, Bernstein, and Mnookin, *The New Wigmore* § 2.1.3 at 40 (cited in note 136).

²⁰⁴ *Id.* at § 2.1.3 at 41 (noting that “modern evidence codifications deem even marginally helpful expert testimony potentially admissible”).

²⁰⁵ See Brandon L. Garrett, *The Substance of False Confessions*, 62 *Stan L Rev* 1051, 1062-66 (2010).

false confessions, if it otherwise satisfies the requirements of admissibility, can appreciably assist jurors in understanding not simply that the phenomenon exists, but also the variables associated with its occurring.

A second function for framework evidence that might assist jurors is to provide the general scientific basis on which later diagnostic judgments are made. This function might be particularly necessary when the diagnostic expert is not familiar with the underlying science or technology that he or she is applying in the particular case, a situation that frequently occurs in forensic and medical contexts. For example, a DNA or fingerprint identification that is based on contemporary testing techniques is likely to be carried out by a technician who is generally unfamiliar with the basic science underlying the technology, often does not have an advanced degree, and is following a protocol created by others.²⁰⁶ Courts might expect, and parties might insist, on having a framework expert testify to the general science relating to a diagnostic test. Similarly, framework experts might provide added value about the efficacy of a range of other diagnostic tests, such as IQ tests employed by psychologists, polygraph tests used by police, and fMRI machines used by neuroscientists. Whether those diagnostic tests are themselves needed to assist jurors is a separate inquiry.

2. Diagnostic testimony.

The measure of whether diagnostic testimony provides added value—in other words, is helpful—in a case depends on an expert's ability to help jurors reason from a valid empirical framework to a valid diagnostic judgment. This is the goal, for example, of testimony typically provided by clinicians: doctors help jurors decide whether the ingestion of a drug known to be associated with cancer caused cancer in the case at hand, and psychiatrists assist jurors in deciding whether the research on schizophrenia's effect on reasoning applies to the defendant on trial. Sometimes this testimony is of suspect validity. Even if it is not, it must still add value.

The situation in which this value is least likely to be present is when the framework evidence is such a good fit with the case

²⁰⁶ See David H. Kaye and George Sensabaugh, *Reference Guide on DNA Identification*, in *Reference Manual on Scientific Evidence* 129, 134 (National Academies Press 3d ed 2011).

that reasonably intelligent jurors need no further assistance applying the empirical framework to the case at hand. Consider, for instance, a case involving whether a battered woman acted reasonably in killing her husband rather than leaving the battering relationship. If the relevant framework research indicated that one of the factors that contributes to a failure to leave the relationship is PTSD, a qualified psychiatrist could provide considerable help to jurors by offering a diagnostic opinion on whether the defendant suffered from PTSD, assuming the professional carried out a proper evaluation.²⁰⁷ If, on the other hand, the factors that contribute to the failure to leave are typically nonclinical—for instance, fear for children left behind, lack of financial resources, and a lack of safe alternatives, such as shelters or nearby family—then a diagnostic expert might not be needed.²⁰⁸ Presumably, a jury could determine, as well as any expert, whether the defendant might have failed to escape the abuse for one or more of these reasons. An opinion by an expert on this issue would probably be akin to an argument the attorney makes during closing and thus would not be helpful.

E. Exclusion on Prejudice Grounds

Federal Rule of Evidence 403 and similar state rules give trial courts the authority to “exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.”²⁰⁹ Rule 403, therefore, gives courts considerable discretion to exclude, or possibly limit the scope of, otherwise-admissible scientific evidence, for any of the reasons listed. In the following discussion we collect these reasons under the rubric of prejudice.

1. Framework testimony.

Some types of framework evidence might be so complicated that juries will be unable to fathom them or so duplicative or obvious that they constitute a waste of time. But the primary danger of

²⁰⁷ See Faigman, et al, 2 *Modern Scientific Evidence* at 304–06 (cited in note 137) (describing research on association of PTSD and battered women).

²⁰⁸ See id at 314–15 (describing research on the correlation between these types of factors and a battered woman’s failure to leave a battering relationship).

²⁰⁹ FRE 403.

framework evidence, as far as the courts are concerned, is not that the jury will misunderstand it or disregard it but rather that it will give the evidence too much weight. For instance, some courts have worried that testimony about the foibles of perception and memory will lead juries to distrust all eyewitnesses.²¹⁰ Courts have expressed similar concerns about framework evidence relating to false confessions,²¹¹ syndromes,²¹² and profiles.²¹³ In essence, the courts decide that the relatively weak fit of framework evidence is outweighed by its prejudicial impact.

If, however, framework evidence is material and scientifically valid, and if it adds value to what the jury could figure out for itself, it should usually be admissible. Rule 403 permits exclusion on prejudice grounds only if the prejudicial impact of the jury would “substantially outweigh” its probative value.²¹⁴ Especially when cross-examination is available to point out that general scientific information does not necessarily dispose of the individual case, the fears expressed by these courts seem unfounded.

2. Diagnostic testimony.

The potential for diagnostic testimony to confuse, distract or overinfluence the jury is best explored by examining the level of inference a diagnostic expert should be allowed to reach in explaining his or her opinion to the jury. In some cases the leap from observed facts (striations on a bullet, fingerprints on a glass, scar tissue on a body) to final conclusions (the defendant’s gun fired the bullet, the fingerprints are the defendant’s, the victim’s injury came from a knife) is relatively small. In these situations,

²¹⁰ See, for example, *People v Enis*, 564 NE2d 1155, 1165 (Ill 1990) (upholding the trial court’s exclusion of testimony on eyewitness identification because “[i]t would be inappropriate for a jury to conclude, based on expert testimony, that all eyewitness testimony is unreliable”).

²¹¹ See *Commonwealth v Harrell*, 65 A3d 420, 431 (Pa Super 2013) (excluding framework testimony about false confessions because “the testimony could confuse the issue by suggesting causal relationships which are not borne out by the research actually conducted”).

²¹² See *Commonwealth v Seese*, 517 A2d 920, 922 (Pa 1986) (excluding testimony about the characteristics of a sexual abuse victim because it “would encourage jurors to shift their focus from determining the credibility of the particular witness who testified at trial, allowing them instead to defer to the so-called ‘expert’ assessment of the truthfulness of the class of people of which the particular witness is a member.”).

²¹³ See *Flanagan v State*, 625 S2d 827, 828–29 (Fla 1993) (holding testimony based on “sex offender profile” inadmissible in part because “[e]stablishing that a defendant has a certain character trait in order to show he acted in conformity with that trait on a certain occasion is forbidden by the rules of evidence”).

²¹⁴ FRE 403.

assuming the other prerequisites for expertise are met (validity, and so on), the prejudice inquiry is relatively simple: will the jury be able to independently assess the validity of the expert's conclusions or will the expert's credentials lull the jury into passive acceptance of whatever the expert says? The court will have to balance the latter possibility with the likelihood that if the diagnostic expert is not permitted to state his or her conclusions, but rather is confined to reporting observed facts, the testimony will be rendered much less useful to the untrained jury. In such cases, the helpfulness of the testimony is closely associated with the expert's ultimate conclusions.

In other types of cases—involving, for instance, psychiatric testimony, economic analysis, or opinions about how an accident occurred—the connection between observed facts and final conclusions is more attenuated, and prejudice analysis thus becomes more complicated. For instance, in *Clark v Arizona*,²¹⁵ the Supreme Court refused to strike down an Arizona statute that prohibited not only expert testimony focused on a defendant's capacity to form mens rea (what might be called the ultimate issue in a case in which the issue is whether the defendant had mens rea), but also expert testimony about the defendant's diagnosis, or what the Court called “mental-disease” evidence (which falls far short of the ultimate issue).²¹⁶ The Court reached this conclusion, it explained, because of the “controversial character of some categories of mental disease,” and because of “the potential of mental-disease evidence to mislead.”²¹⁷ Put in evidentiary terms, the majority in *Clark* was concerned that the fact finder might give diagnostic testimony about the defendant's mental disorder undeserved weight on the related, but different, mens rea issue, and thus appeared to believe that prejudicial impact of such testimony could easily outweigh its probative value.

The issue raised in *Clark* can arise in any case involving diagnostic testimony, because, compared to framework testimony, diagnostic testimony is more likely to be suspect in terms of validity, at the same time its particularized nature and better “fit” is more likely to influence the fact finder's determination in the

²¹⁵ 548 US 735 (2006).

²¹⁶ *Id* at 774–76 (concluding that it is reasonable to channel “mental-disease and capacity evidence” to the insanity issue, in which the defendant has the burden of persuasion).

²¹⁷ *Id* at 774.

case at hand. The usual response to this concern is that the adversarial process—cross-examination, rebuttal experts, etc.—can correct for this problem.²¹⁸ However, the majority in *Clark* was not swayed by the dissent's observation that "[w]e have always trusted juries to sort through complex facts in various areas of law."²¹⁹

Whatever the correct outcome in *Clark* itself,²²⁰ the general concern the *Clark* majority raises is worth taking seriously. Diagnostic experts are, by definition, extrapolating from general scientific principles. At some point, even if their opinions are based on vast experience and accepted protocols, they can reach too far. When diagnostic testimony is at issue, the prejudice inquiry is a crucial tool for ensuring that expert testimony aids the adjudication process. This inquiry requires close attention to whether the testimony rests on a valid empirical framework that permits extrapolation to an individual case, as well as the extent to which the testimony departs from the diagnostic skills the expert is known to possess (as a result of proficiency testing or some other measure of validity). The less supportive the framework or the greater the departure from a validated diagnostic skill set, the more attentive the court must be to the ability of the adversarial system to expose for the jury any flaws in the testimony.

CONCLUSION: INTEGRATING BEST PRACTICES INTO JUDICIAL ADMISSIBILITY JUDGMENTS

As the foregoing discussion indicates, and as *Daubert* and its progeny emphasized, analysis of when witnesses should be allowed to testify as either framework or diagnostic experts ultimately involves balancing many factors. Courts need to consider the extent to which the testimony:

- (1) is material to issues in the case, that is, can be said to fit the case both legally (as mandated by the relevant substantive

²¹⁸ See, for example, *Barefoot v Estelle*, 463 US 880, 901 (1983) (stating that "the adversary process" can "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").

²¹⁹ *Clark*, 548 US at 793 (Kennedy dissenting), quoting *United States v Booker*, 543 US 220, 289 (2005) (Stevens dissenting).

²²⁰ See Christopher Slobogin, *The Supreme Court's Recent Criminal Mental Health Cases*, 22 *Crim Just Magazine* 8, 12 (2007) (arguing that *Clark* "exaggerates the gullibility of judges and juries about psychiatric evidence; research shows that laypeople are very skeptical about such evidence, especially when it is presented by the defense").

- law) and empirically (as indicated by an assessment of external validity);
- (2) is proffered by a witness who is qualified in the relevant field;
 - (3) evidences internal validity, which can be measured in terms of the extent to which the basis of the testimony:
 - (a) has been subject to testing or some other verification process; (b) can be analyzed using error rates; (c) was developed in accordance with adequate standards; (d) has been exposed to peer review; (e) is generally accepted by knowledgeable experts in related, scientifically-oriented fields; and (f) is the product of the type of rigorous analysis expected of experts in the field;
 - (4) adds value (in other words, is helpful), because it provides useful information to the fact finder; and
 - (5) avoids misleading or distracting the fact finder.

Consistent with *Daubert's* cautious approach,²²¹ these considerations are intended to be guidelines, rather than strict criteria or a formal checklist that can be employed in a mechanical fashion. The admissibility decision for both framework and diagnostic evidence is a matter of judgment and these considerations are intended as guides to that judgment, not substitutes for it.

Nonetheless, these guidelines should be seen as a set of judicial best practices that govern admissibility analysis in three substantive ways. First, it should be clear from the foregoing that proffered expert testimony must meet *all five* of these factors in order to be admitted. Failure on any one of them should lead to exclusion. Immaterial evidence is inadmissible regardless of its validity. Invalid evidence is inadmissible even if it appears to add value to the case. And so on.

At the same time, as captured by the balancing metaphor, in those situations in which the threshold for each factor is met, a weakness in one can be compensated for by strength in the others. For instance, expert evidence that does not have perfect empirical fit (because, for instance, the validation sample used for the research did not contain people sufficiently similar to the parties in the instant case) may still be admissible if it has high indicia of validity; as scientists have long recognized, tradeoffs

²²¹ *Daubert*, 509 US at 594 (“The inquiry envisioned by Rule 702 is, we emphasize, a flexible one.”).

between external and internal validity may (and probably must) be made.²²² Similarly, as we have pointed out at several different points in this Article, weaknesses in connection with any one of the indicia of internal validity may not be dispositive if other indicia can pick up the slack. However, “convergent validity”—in which most or all of the factors indicate the basis of the testimony is trustworthy—is obviously preferable and is more likely to permit the internal-validity factor to make up for deficiencies in the other factors.

A third aspect of these best-practice guidelines is that the five factors they identify must be balanced separately—and differently—depending on whether the expert is offering framework evidence or diagnostic evidence. That is the crucial (and new) point of this Article. Before summarizing these differences and providing an example of how they might apply, we briefly turn to the predicate issue of distinguishing between general framework evidence and particularized diagnostic evidence.

A. Distinguishing Framework and Diagnostic Evidence

Differentiating between framework and diagnostic evidence is fundamental to understanding the use of science in the courtroom. Most of the time, the distinction will be straightforward and uncontroversial. Clearly, for instance, framework evidence is being offered by the psychologist who relies on the scientific literature in testifying that certain factors tend to produce inaccurate eyewitness identifications or false confessions, or an epidemiologist who testifies, based on the research, that certain drugs are possible causes of cancer. Just as clearly, an expert who offers to testify that the plaintiff’s birth defects were caused by the drug in question, or that the plaintiff’s tire was defectively manufactured, is offering diagnostic expert testimony, even if the expert also references general scientific research in the course of doing so. Returning to the breast-cancer litigation example we used at the beginning of this Article, the difference is between an expert who says, “When factors A, B, and C are present, all else held constant, women who do not get a mammogram by age forty have an X percent chance of getting cancer in the next five years” (a framework statement) and an expert who

²²² See Monahan and Walker, *Social Science in Law* at 69 (cited in note 105) (“Often a research strategy that yields results high in internal validity does so at the cost of leaving external validity questions unanswered, and vice versa.”).

says, "Given what I know about Ms. Jones, including factors A, B, and C, she had a Y percent chance of getting cancer before age forty" (a diagnostic statement).

But some types of expert testimony are harder to classify. Consider an expert who testifies in response to detailed hypothetical questions that replicate the facts of the case, but who has not performed any evaluation of the relevant party and has not independently heard any evidence in the case. Is this expert a diagnostic expert or a framework expert? Presumably the witness is relying on research in answering the hypothetical question, but he or she is also providing very case-specific information based on that research, albeit without the usual evaluation process. When the hypothetical permits the expert to drill down into the details of the particular case, and is so tailored to it that, in effect, only the party's name is missing, the testimony is diagnostic. But less detailed hypothetical questioning can be more difficult to classify.²²³

The ultimate question, of course, is whether the answers to the hypothetical questions should be admissible. We believe the analysis depends on how the expert responds to the question. If the expert says something to the effect of, "On those facts, the individual would be likely to [be insane/give inaccurate testimony/be diagnosed with cancer]," the testimony would be inadmissible, unless the expert can produce scientific evidence backing up the claim that on this set of facts the assertion is "likely" (which in most cases will not be possible). Nor should the expert be able to hide behind meaningless assertions of "reasonable [medical/psychological/scientific] certainty" in answering such questions.²²⁴ Assume, instead, however, that the expert answers

²²³ Compare *Tillman v State*, 354 SW3d 425, 428–32 (Tex Crim App 2011) (allowing expert on eyewitness testimony to answer a string of hypothetical questions regarding: the effect of having the defendant be the only person in a lineup who was also in a previous photo array, the effect of pointing out the defendant on the street to an eyewitness and then putting the defendant in a lineup, the effect of working with a police artist, and the effect of putting the suspect in a lineup when he is the only one with facial hair), with *State v Yusuf*, 800 A2d 590, 607 n 12 (Conn App 2002) (allowing use of extremely detailed hypothetical questions during direct examination of expert on battered woman syndrome, including the number of assaults and how and when they occurred, that replicated what the defendant did to the victim). We think *Tillman* involves framework testimony and *Yusuf* involves diagnostic testimony.

²²⁴ See Lewin, 57 Md L Rev at 502 (cited in note 71), concluding that the phrase "reasonable medical certainty":

was incorporated into legal doctrine not because it best served certain instrumental purposes, but by virtue of the judiciary's uncritical acceptance of

the hypothetical in the following terms: "Based on the research, and all else being equal, the individual in the hypothetical would be more likely to [be insane/be inaccurate/be diagnosed with cancer] than when the stated facts are not present." This more modest statement sets the response within an empirical framework that is more likely to have scientific backing.

B. The Best-Practice Guidelines Applied

Once the distinction between framework and diagnostic evidence has been made, the court must apply the five guidelines we outlined earlier in deciding whether the evidence is admissible. The factors operate very differently depending on the level of expert testimony being proffered. These differences, detailed throughout this Article, are summarized here for each of the factors.

Relevance: By definition, diagnostic testimony is more likely to be material than framework testimony. Individualized testimony will always have a better legal fit with the facts than testimony based on general research findings. When it comes to empirical fit, however, diagnostic testimony that reaches beyond the expert's customary practice may well be irrelevant, while framework evidence's relevance will vary widely depending upon the underlying research's external validity.

Qualifications: A qualified framework expert will need to be knowledgeable about the research literature, but need not possess evaluation experience or skills. Conversely, a minimum degree of experience and requisite skill will usually be obligatory for a diagnostic expert. At the same time, in contrast to the framework expert, the latter type of expert need not always be immersed in the background science (as with DNA technicians).

Validity: Because of its explicitly scientific nature, framework evidence should be subject to the most demanding internal-validity tests. Courts should require proof of a rigorous verification process and an explanation for the absence of error rate information. Some diagnostic expertise is subject to verification as well, through proficiency examinations or other feedback loops. But more often courts evaluating the scientific validity of diagnostic testimony will need to look at how peer experts

a prevailing usage. Having no intrinsic meaning, the phrase was not especially well-suited to its eventual dual roles as a standard of admissibility and a standard of proof.

customarily function, with a focus on whether standard practices have been developed by knowledgeable experts in the field for dealing with the precise issue at stake in the case. In other words, *Daubert's* factors, and in particular its first three, are crucial considerations in evaluating the admissibility of framework evidence, whereas *Frye's* focus on well-accepted protocols and customary practice, with the caveats we have expressed in this Article, will often be more useful in assessing the admissibility of diagnostic expert evidence.

Added Value (Helpfulness): Both framework and diagnostic testimony will most often be helpful when they educate the jury about complicated or arcane subjects, or when they challenge common misconceptions. But the more closely framework evidence fits with the facts of the case, the less helpful diagnostic testimony may be, because the fact finder will more likely be able to draw its own conclusions from what it has learned via the framework evidence. In other words, when framework evidence is a good fit to the facts of the case, the fact finder, guided by lawyers and judicial instructions, may not need the diagnostic expert: the diagnostic expert will add no value to what the framework expert has already testified to.

Prejudice: Framework evidence is most likely to be prejudicial in the sense that it overcorrects (by, for instance, leading a jury to distrust all eyewitness testimony). Diagnostic evidence is most likely to be prejudicial the further it departs from what framework scientists can say with a high degree of certainty. In both cases, courts will need to assess the extent to which the adversarial process can minimize these dangers, but the presumption should be that it will usually be able to do so if analysis of the other factors point in the direction of admissibility.

To illustrate how these differences would play out in an actual case, consider an example involving expert evidence about eyewitness testimony in a criminal case. Suppose first that the defendant offers expert framework testimony that cross-racial identifications are less accurate on average than same-race identifications. Regarding *relevance*, this testimony legally fits if the identification involved in the case was cross-racial. Its empirical fit is less obvious, since most of the research in this area is laboratory research.²²⁵ Yet there are good reasons to believe

²²⁵ See Christian A. Meissner and John C. Brigham, *Thirty Years of Research on the Own-Race Bias in Memory for Faces: A Meta-analytic Review*, 7 *Psychology Pub Pol & L* 3, 13 (2001) (describing 39 research studies, all of which involved laboratory methodology).

that the laboratory research of this sort should generalize to the wider world.²²⁶ The expert's *qualifications* should depend on his or her knowledge of the research literature. Because that literature is voluminous, some experts may not be familiar with its nuances and, for that reason, might be rejected at this stage, but those experts who have been adequately trained should be qualified to testify. Turning to the various scientific validity factors, given the voluminous literature just mentioned the *testability* criterion is satisfied, since the subject of cross-racial identifications has been extensively examined by many different teams of researchers.²²⁷ The possible *error*, including the variability and limited effect sizes, of this research are also well known.²²⁸ Most of the research that an expert in this area would rely upon is published in well-regarded *peer-reviewed* journals and those in the field of cognitive psychology consider its findings to be *well accepted* and the result of *standard* and *rigorous* methodology.²²⁹ Next to consider is the *helpfulness* factor. While some judges have doubted the added value of this type of testimony, increasingly courts have come to appreciate jurors' lack of appreciation for the limits of eyewitness identifications.²³⁰ Finally, *prejudice*, in particular the possibility that jurors will dismiss all cross-racial identifications out of hand,²³¹ might augur against admissibility. But this factor is probably not substantial enough to lead to exclusion of otherwise admissible evidence if the adversarial process functions as it should. Taken together, then, the five factors of relevance, qualifications, validity, added value,

²²⁶ *Id.* at 27 ("Overall, the [own-race-bias phenomenon] was found to be a reliable and generalizable phenomenon . . ."). See also generally Dawn E. McQuiston and Roy S. Malpass, *Validity of the Mock Witness Paradigm: Testing the Assumptions*, 26 L & Hum Behav 439 (2002) (describing research suggesting good external validity using mock witnesses). But see Deborah Bartolomey, *Cross-Racial Identification Testimony and What Not to Do About It: A Comment on the Cross-Racial Jury Charge and Cross-Racial Expert Identification Testimony*, 7 Psychology Pub Pol & L 247, 249 (2001) (arguing that the research does not translate to real-world cases).

²²⁷ See Meissner and Brigham, 7 Psychology Pub Pol & L at 21 (cited in note 225) (discussing the results in thirty-nine research studies on cross-racial identification).

²²⁸ See *id.* at 15–21 (describing a metareview summarizing overall effect sizes).

²²⁹ See *id.* at 23 (describing studies from peer-reviewed journals).

²³⁰ See *State v Clopten*, 223 P3d 1103, 1113 (Utah 2009) (noting that "trial judges have often excluded eyewitness experts on grounds that the testimony will not be helpful to the jury," but concluding, to the contrary, that "[a]s the research makes clear, the topics covered by eyewitness experts are often beyond the common knowledge of ordinary jurors and usually cannot be effectively elicited through cross-examination alone").

²³¹ See Bartolomey, 7 Psychology Pub Pol & L at 252 (cited in note 226) ("Identification experts will not help the jurors reach a rational verdict, but will muddle the issues, invade the jury's province, and escalate the already considerable expense of criminal trials.").

and prejudice largely favor admitting framework testimony on cross-racial identifications in the ordinary case.

Suppose, however, that the eyewitness expert also offers a diagnostic opinion on the accuracy of the witness's cross-racial identification. The same five factors now point strongly in the opposite direction. Although, by definition, such an opinion would legally fit the case at hand, nothing in the expert's practice provides a foundation on which empirical fit could be measured. For related reasons, a witness could not have received training in such clinical assessments, and so should not be qualified to offer a diagnostic opinion on the issue. With respect to both the qualification and validity issue, the claim might be made that an eyewitness expert who has testified in large numbers of cases involving cross-racial identifications possesses the "wisdom" to pinpoint when a witness who has made such an identification is likely wrong. But experience cannot qualify as a test of an expert's diagnostic judgment in such cases; tellingly, neither a protocol for making these judgments nor a feedback loop that tells experts whether their experience produces accurate judgments exists. In short, diagnostic assessments are not customary in this particular field (in contrast, for instance, to customary practice in the medical profession).²³² Moreover, the added-value criterion strongly weighs against eyewitness diagnostic testimony, since once jurors hear the framework evidence they can be expected to handle the inferential challenges

²³² Our position here is thus directly contrary to that of Professors Susan T. Fiske and Eugene Borgida, who are prominent proponents of allowing framework experts to offer opinions on individual group members. They argue that the custom of limiting such testimony to physicians and other diagnostic experts is inapt:

A sharp distinction between general and specific causation is alien to psychologists and most scientists. The ability to generalize to specific circumstances is probabilistic, and applying knowledge in court is not different in principle from applying scientific evidence in other contexts. However, the confidence with which experts can generalize varies, depending on the state of available, relevant science. *Like physicians' ability to diagnose a patient's physical symptoms*, scientists' ability to link general and specific causation for the individual case crucially depends on the quality of available scientific evidence and the scientist's relevant expertise. Qualified social scientists who provide general, relevant knowledge and apply ordinary scientific reasoning may offer informed opinion about the individual case, but probabilistically. None of this usurps the triers of fact of their role, as they are capable of drawing their own conclusions, with scientific judgments as one input.

Susan T. Fiske and Eugene Borgida, *Standards for Using Social Psychological Evidence in Employment Discrimination Cases*, 83 Temple L Rev 867, 875-76 (2011) (emphasis added; emphasis omitted).

of reasoning from the group data to the individual case²³³ (a capacity that can again be contrasted to the typical medical-causation case, in which jurors might have considerable difficulty reasoning from multiple types of studies in epidemiology and toxicology to the individual case). Finally, the prejudice associated with eyewitness diagnostic testimony, in the sense that it could distract a jury from the scientific uncertainty inherent in the field, probably substantially outweighs its limited probative value.

Differential application of the five factors gleaned from *Daubert* of the type illustrated by this example is crucial if courts are to succeed at balancing the numerous legal and scientific considerations that influence when general research may be heard in court and the extent to which experts may apply that research to help resolve specific cases. Only in this way can the distinctions between framework and diagnostic evidence be openly acknowledged and confronted. Unless courts develop and refine evidentiary best practices akin to those proposed here, they will not be able to make full use of the knowledge that science creates and will at the same time heighten the risk that value-laden but empirically valueless diagnostic speculation will undermine the adjudication process.

²³³ See *Clopton*, 223 P3d at 1114 (“As long as the expert does not attempt to tell the jury that a specific eyewitness identification either is or is not accurate, then the expert has not impinged on the jury’s duty as the sole evaluator of witness credibility.”).