Toward a Jurisprudence of Psychiatric Evidence: Examining the Challenges of Reasoning from Group Data in Psychiatry to Individual Decisions in the Law

CARL E. FISHER,* DAVID L. FAIGMAN** & PAUL S. APPELBAUM***

Psychiatry is an applied science. It thus shares a characteristic of all applied science in that it is ultimately applied at two levels—general and specific. Scientific research inevitably focuses on aggregate data and seeks to generalize findings across persons, places, or things. However, in the courtroom, as is true in other applied settings, the focus is usually on an individual case. Thus, psychiatry presents the challenge inherent in all scientific evidence of reasoning from group data to an individual case, which is termed the "G2i problem." Psychiatry, unlike many scientific fields that come to court, also confronts the G2i problem in its daily practice since mental health professionals routinely diagnose and treat individuals based on aggregate data. Yet approaches to the G2i problem in clinical psychiatry do not necessarily comport with the ways in which aggregate data is applied to an individual case in the courtroom.

In this Article, we employ the G2i lens to examine the admissibility of psychiatric expert testimony in regards to both general research findings, or "framework evidence," and application of those general findings to specific cases, or "diagnostic evidence." Although the

* Assistant Professor of Clinical Psychiatry, Department of Psychiatry at Columbia University and New York State Psychiatric Institute, New York, NY.

** John F. Digardi Distinguished Professor of Law, University of California Hastings College of the Law; Co-Director, UCSF/UC Hastings Consortium on Law, Science & Health Policy; Professor, Department of Psychiatry, University of California San Francisco School of Medicine.

*** Elizabeth K. Dollard Professor of Psychiatry, Medicine and Law, and the Director of the Division of Law, Ethics & Psychiatry in the Department of Psychiatry at Columbia University and New York State Psychiatric Institute, New York, NY.

We thank the members of the MacArthur Foundation's G2i Committee, who provided valuable comments and suggestions throughout the development of this Article, including Josh Buckholtz, Andre Davis, Philip Dawid, Nancy Gertner, Peter Imrey, John Monahan, Russell Poldrack, Marcus Raichle, Valerie Reyna, Christopher Slobogin, and Tracey Woodruff. We are also indebted to members of the MacArthur Law and Science Network for their helpful comments on an earlier draft, with special thanks to Richard Bonnie, Owen Jones, and Stephen Morse. Preparation of this Article was supported, in part, by a grant from the John D. and Catherine T. MacArthur Foundation. 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.

rules of evidence that apply to "G" and to "i" are the same, the scientific and professional considerations with which each must be evaluated are fundamentally different. G2i inferences provide a useful lens by which the interactions of psychiatry and law can be better understood and managed.

I. Introduction: Reasoning from Group Data in Science to Individual Decisions in the Law (the "G2i Problem") 687
II. Where Law and Psychiatry Intersect 690
III. G2i in Law and Psychiatry 693
A. *G2i in Law: Admissibility Standards* 696
1. Common Requirements of "Qualifications," "Relevance," and Helpfulness 697
2. Measuring Reliability and Validity 698
i. *Frye* and Tests of "Acceptance" 699
ii. *Daubert* and Tests of "Validity" 700
B. *G2i in Psychiatry* 702
1. *Diagnostic* Frameworks 702
2. Functional Assessment 709
IV. Psychiatry in the Courtroom: A User's Guide 712
A. *Current Approaches to Admissibility* 713
B. *G2i Analysis of Admissibility* 716
1. Relevance or "Fit" 717
i. Framework Evidence on "Fit" 717
ii. *Diagnostic* Evidence on "Fit" 720
2. Helpfulness (i.e., "Assist the Trier of Fact") 721
i. Framework Evidence on Helpfulness 722
ii. *Diagnostic* Evidence on Helpfulness 722
3. Reliability and Validity of Methods and Principles 723
i. Testing 724
a. Framework Evidence on Testing 724
b. *Diagnostic* Evidence on Testing 728
ii. Error Rate 730
a. Framework Evidence on Error Rate 731
b. *Diagnostic* Evidence on Error Rate 733
iii. Peer Review and Publication 736
a. Framework Evidence on Peer Review and Publication 737
b. *Diagnostic* Evidence on Peer Review and Publication 739
iv. "Widespread Acceptance" 740
a. Framework Evidence on "Widespread Acceptance" 740
b. *Diagnostic* Evidence on "Widespread Acceptance" 742
v. "Same Level of Intellectual Rigor" 743
a. Framework Evidence on "Intellectual Rigor" 743
b. *Diagnostic* Evidence on "Intellectual Rigor" 744
4. Unfair Prejudice 746
i. Framework Evidence on Unfair Prejudice 746
ii. *Diagnostic* Evidence on Unfair Prejudice 747
C. *G2i and the Evolution of Psychiatric Expert Testimony* 748
1. Dimensional Approaches to Diagnosis 748
2. Neuroscience-Based Testimony 750
V. Conclusion 752

I. Introduction: Reasoning from Group Data in Science to Individual Decisions in the Law (the "G2i Problem")

In the conventional view, scientific fields advance through the concerted efforts of researchers dedicated to studying phenomena to better describe, predict, and not infrequently, control them. As basic research data accumulate, they often are applied to specific instances of the phenomena being studied. Meteorologists, for example, study weather systems to identify variables that will permit accurate forecasting of rain on a particular day, and geophysicists study earthquakes to anticipate when the plates along a given fault line will shift. Although the former have so far been more successful than the latter, the basic framework they employ is fundamentally similar: gather data from enough instances of the phenomenon in question so that in the aggregate they provide sufficient understanding to permit description, prediction, and (sometimes) intervention. In fact, this basic structure of scientific inquiry is consistent across the many disciplines interested in applying their data to real-world events, including engineering, medicine, economics—and, of course, the subject of this Article, psychiatry.[1]

Despite the pervasiveness of the view described above, the scientific enterprise is not principally designed to provide categorical assessments of particular cases. Indeed, the focus of science is the analysis of aggregate data, and the language of science is statistics. Whether a scientist is studying violent storms or violent people, research findings are expressed in terms of likelihoods and probabilities. Meteorologists might be able to specify the likelihood that storms of the type bearing down on a particular city will evolve into a hurricane, and psychiatrists may be able to specify the likelihood that people who are similar in certain ways to a respondent in a civil commitment hearing will be violent in the future. But these are probabilistic assessments that can only inform—not determine with certainty—whether a city will be hit by a hurricane or whether a respondent will be violent. The degree of uncertainty becomes the operative question when science is applied to particular settings. In the forecasted hurricane example, the decision of whether to board up windows, ride out the storm, or evacuate the city will depend in part on the uncertainty of the forecast.[2] When it comes to the prediction of violence, the decision to put a person who is mentally ill on a 72-hour psychiatric hold should similarly take into account the uncertainty of the prediction.

1. In this Article, we use "psychiatry" as shorthand for the mental health professions as a whole. Many of the evidentiary functions discussed are also performed by other mental health disciplines, especially psychology.

2. This uncertainty can take many forms, including location, duration, and severity.

When science is used to make decisions about individual cases, it often requires categorical judgments to be made in light of the uncertainties of the general data. In medicine, treatment decisions are made after considering research on the success rates of possible treatments, which may differ between studies due to as-yet-unspecified variables. In engineering, particular buildings may be closed based on industry research on the likelihood of structural failures, but different studies may similarly reflect varying data. In conservation biology, animals are put on endangered species lists based on research on extinction risks, which are difficult to quantify. Psychiatrists operate similarly, having to digest general data about groups of patients and make judgments about medications, therapy, and commitment for individuals. These judgments are very often, but not always, informed by diagnoses and conclusions about patients' functional abilities, which are in turn informed and defined by aggregate data from systematic studies or clinical experiences.

The law makes similar demands on the applied sciences. Indeed, in the courtroom, science is regularly employed to inform categorical decisions about individuals. As Faigman, Monahan, and Slobogin note, "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."[3] Was the eyewitness, who saw the perpetrator for a few seconds and who made a cross-racial identification of the defendant, accurate? Did the defendant's toxic product cause the plaintiff's illness? Is the defendant so intellectually disabled that he cannot be executed under the standards prescribed by the Eighth Amendment?[4] The list of examples is nearly endless and has evoked a term of art to describe the dilemma of applying group data to individual instances, "the G2i problem."[5]

The challenges for a scientific expert in responding to these questions can be illustrated by considering the example of eyewitness identification. Studies of groups of people have demonstrated that—in addition to the problems that attend identification of suspects in general—cross-racial identification is particularly fraught with error.[6] Most people, most of the time, are less adept at identifying persons of another

3. David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417, 420 (2014).

4. U.S. CONST. amend. VIII ("Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted.").

5. Faigman, Monahan & Slobogin, *supra* note 3, at 419–20.

6. *See* Gary L. Wells, Amina Memon & Steven D. Penrod, *Eyewitness Evidence: Improving Its Probative Value*, 7 PSYCHOL. SCI. PUB. INT., no. 2, 2006, at 45, 52 (reporting that "the chance of mistaken identification is 1.56 times greater in other-race than in same-race conditions").

race than identifying persons of their own race.[7] But not every attempted identification of a person of a different race is inaccurate. The tendency to err—though demonstrable at the level of decisions made by many people trying to identify persons from other races—is not absolute. Many cross-racial identifications are, in fact, accurate. How then can an expert apply the group data that suggest a greater tendency toward error in cross-racial cases to the categorical determination that needs to be made in a particular case (i.e., whether this particular witness erred)?[8] Even more to the point, should courts permit eyewitness experts to opine on the likely accuracy of a particular identification? And, if eyewitness experts are not given license to comment on particular applications of their data, why should any other expert be allowed to do so?

It is readily apparent that psychiatric testimony in court is subject to exactly this dilemma. As a discipline, psychiatry is oriented toward digesting large amounts of information for the purpose of categorizing individuals as falling within particular diagnoses, suffering certain conditions, or manifesting certain impairments. At the same time, however, the eyewitness example is also inapposite to the situation of psychiatry. Psychiatrists, unlike eyewitness researchers, customarily apply general research findings to particular cases. In that sense, psychiatry is more similar to other areas of medicine or even meteorology than it is to the typical research science.

Yet, the professional judgments involved in reasoning from the general to the individual case in ordinary psychiatric practice are not necessarily coterminous with the judgments required in court. Whereas psychiatrists assess individual patients for purposes of making individual treatment decisions, courts are interested in assessing the psychological characteristics of individuals to aid legal decisionmakers in dispensing fair and just outcomes pursuant to applicable law. For psychiatric testimony to aid in this process, psychiatric experts must apply to individual defendants or litigants data that are derived from the study of groups in a valid manner, and judges must understand the predicates

---

7. Christian A. Meissner & John C. Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review*, 7 PSYCHOL. PUB. POL'Y & L., no. 1, 2001, at 3, 15.

8. In several areas of expertise (e.g., eyewitness identification, medical causation, employment discrimination, and physical forensic identification), courts have not been consistent in choosing the level of specificity to which they permit or require experts to testify. The basic choice can be conceptualized as between "framework" evidence alone (i.e., describing general scientific propositions, such as a greater tendency toward eyewitness error under certain conditions) and particularized "diagnostic" evidence (i.e., applying the general propositions to individual cases). For a detailed review, see Faigman, Monahan & Slobogin, *supra* note 3, at 424, 432.

for the appropriate use of such data. We suspect that often neither of these *desiderata* are met when psychiatric testimony is introduced.

II. Where Law and Psychiatry Intersect

Law, in both scholarship and practice, has a long history of skepticism about psychiatry. This distrust exists even at the level of diagnostic description, arguably one of the foundational elements of psychiatric evidence. The Supreme Court of the United States has repeatedly held that the law is not bound by psychiatry's official nosology because the usual professional diagnoses of mental disorders are often disputed and potentially inexact.[9] At the statutory level, many mental health laws contain a disorder criterion (i.e., a predicate requirement for an official diagnosis before further arguments about mental health are entertained),[10] but some legal scholars have suggested that this criterion should be discarded in favor of solely addressing the legally relevant behavior and function.[11]

More broadly, the application of psychiatric evidence to legal determinations is sometimes questioned. Statutes frequently and categorically exclude testimony about some types of functional impairment (e.g., specific intent or mental states other than insanity), as well as certain psychiatric disorders, including disorders that have achieved widespread acceptance in clinical practice.[12] Legal attitudes toward psychiatry have been inconsistent over time, however, reflecting shifting views about which diagnoses "count," while the usual clinical concepts of diagnosis and functional impairment have been discarded in some contexts in favor of non-medical concepts such as "mental abnormality."[13] Overall,

9. In *Clark v. Arizona*, 548 U.S. 735, 744 (2006), for example, the Court quoted the fifty-year-old decision of *Greenwood v. United States*: "The only certain thing that can be said about the present state of knowledge and therapy regarding mental disease is that science has not reached finality of judgment." Greenwood v. United States, 350 U.S. 366, 375 (1956). *See also* Kansas v. Hendricks, 521 U.S. 346, 360 n.3 (1997); Addington v. Texas, 441 U.S. 418, 430 (1979).

10. *See infra* notes 122–24 and accompanying text.

11. Stephen J. Morse, *Mental Disorder and Criminal Law*, 101 J. Crim. L. & Criminology 885, 895 (2011) ("Mental disorder per se is not a sufficient criterion for special legal treatment. All mental health laws require further legally relevant behavior, which is in fact the law's primary concern.").

12. *See infra* Part IV.B.2.ii. For an organized list of several exclusions, see Ralph Slovenko, *The Role of Psychiatric Diagnosis in the Law*, 30 J. Psychiatry & L. 421, 427 (2002) (listing statutory exclusions such as sociopathy, personality disorders, and pathological or compulsive gambling). *See also, e.g.*, Ariz. Rev. Stat. Ann. § 13-502(A) (2014) ("A person may be found guilty except insane if at the time of the commission of the criminal act the person was afflicted with a mental disease . . . . Mental disease or defect does not include disorders that result from acute voluntary intoxication or withdrawal from alcohol or drugs, character defects, psychosexual disorders or impulse control disorders.").

13. *See Hendricks*, 521 U.S. at 360 (1997) (diagnosis as a pedophile qualified as a "mental

in both law and society, there is significant doubt regarding psychiatry, especially its ability to distinguish "normal" from "disordered."

Yet, in individual cases as reflected in the ordinary practice of the courts, psychiatric diagnostic categories do matter to the law. The presence of a diagnosis is often a prerequisite for further consideration of evidence about a mental disorder. As such, diagnoses are important components of claims that address the pivotal issues in individual cases, such as culpability, competence, and predictions of future behavior. Beyond this threshold function, however, psychiatric diagnoses can play a number of other roles.[14] Insofar as a given disorder is associated with specific manifestations (e.g., people with schizophrenia are likely to experience delusional thoughts), a diagnosis alerts attorneys and expert evaluators alike to explore the presence and functional impact of such symptoms. Since many disorders have a typical history (e.g., bipolar disorder is marked by intermittent manic or depressive episodes, with substantial recovery of function in between), expert witnesses' retrospective and predictive assessments will often draw on their knowledge of the longitudinal course of particular disorders.[15] Diagnostic formulations allow easier detection of malingered conditions, which are often characterized by atypical clusters of symptoms that are not associated with any known disorder. Finally, diagnoses anchor the clinical judgment of the expert, helping to restrain unfounded speculation that transcends the generally accepted characteristics of the disorder in question. In sum, the process of psychiatric diagnosis is a way for the medical profession to express its best understanding of mental disorders. While there is, as yet, no blood test or brain scan that can definitively establish a psychiatric diagnosis, there is a body of conceptual work and empirical evidence that strengthens the explanatory power of this process.[16]

More broadly, psychiatric diagnosis also serves a useful rhetorical function; it expresses to policymakers and the lay public that some people are "mad" rather than "bad," justifying treatment rather than punish-

---

abnormality" under Kansas' Sexually Violent Predator Act of 1994); Kansas v. Crane, 534 U.S. 407, 413 (2002) ("[T]here must be proof of serious difficulty in controlling behavior. And this, when viewed in light of such features of the case as the nature of the psychiatric diagnosis, and the severity of the mental abnormality itself, must be sufficient to distinguish the dangerous sexual offender . . . from the dangerous but typical recidivist convicted in an ordinary criminal case.").

14. *See* AM. PSYCHIATRIC ASS'N TASK FORCE ON USE & MISUSE OF PSYCHIATRIC DIAGNOSES IN THE COURTS, THE USE OF PSYCHIATRIC DIAGNOSES IN THE LEGAL PROCESS 4 (1991).

15. *See generally, e.g.*, Stanley Zammit et al., *A Longitudinal Study of Premorbid IQ Score and Risk of Developing Schizophrenia, Bipolar Disorder, Severe Depression, and Other Nonaffective Psychoses*, 61 ARCHIVES GEN. PSYCHIATRY 354 (2004), *available at* http://archpsyc.jamanetwork.com/article.aspx?articleid=481989.

16. For discussion of research regarding genetic correlates of mental disorders, see *infra* Part IV.C.2.

ment for those who suffer from mental disorders.[17] By the same token, when the law strays from accepted clinical concepts, such as the invention of a novel "mental abnormality" definition for "sexual predator" laws, the fundamental justifications for confining certain people with alleged mental illnesses can become confused.[18]

Thus, there appears to be a problem: psychiatry is called upon to help the law understand mental disorders, but the law is often dubious and inconsistent when it encounters psychiatric evidence. Certainly, some degree of skepticism is warranted. Psychiatry's own standard reference for mental disorder classification articulates a "Cautionary Statement," and provides that psychiatric diagnoses developed for a clinical context do not necessarily translate to legal settings.[19] There is thus much room for improved clarity at the intersection of psychiatry and law. We suggest that a development from the broader field of science and law might help clarify some of the relevant issues, including when such evidence is admissible.

Recently, legal and scientific scholars have begun to more specifically parse out the process of reasoning from group data to individual cases.[20] The relevance of G2i inference to psychiatry is clear: A diagnostic and functional assessment is essentially an inference that connects

17. *See* Vanessa L. Kolbe, *A Proposed Bar to Transferring Juveniles with Mental Disorders to Criminal Court: Let the Punishment Fit the Culpability*, 40 VA. J. SOC. POL'Y & L. 418, 436 (2007) ("[M]any juvenile offenders have underlying mental disorders, and simply punishing them seems counterintuitive when their mental disorder may have contributed to their criminal behavior."). *But see* Richard C. Boldt, *Rehabilitative Punishment and the Drug Treatment Court Movement*, 76 WASH. U. L.Q. 1205, 1208 (1998) (discussing drug treatment courts, which meld "substance abuse treatment and punishment").

18. Whereas it is fairly clear that a person suffering from a "mental illness" should receive treatment when civilly committed, the non-medical category of "mental abnormality" creates confusion, since the offender has been civilly committed and might be categorized as "suffering" from a condition, but one that is not an "illness" and not treatable. *See, e.g.*, AM. PSYCHIATRIC ASS'N TASK FORCE ON SEXUALLY DANGEROUS OFFENDERS, DANGEROUS SEX OFFENDERS 11–17 (1999) (tracing the history of "sexual psychopath laws," and noting the "purpose" of such laws as shifting from "therapeutic" to "incapacitative"); W. Lawrence Fitch, *Sexual Offender Commitment in the United States: Legislative and Policy Concerns*, 989 ANNALS N.Y. ACAD. SCI. 489, 500 (2003); Sameer P. Sarkar, *From* Hendricks *to* Crane*: The Sexually Violent Predator Trilogy and the Inchoate Jurisprudence of the U.S. Supreme Court*, 31 J. AM. ACAD. PSYCHIATRY & L. 242, 247 (2003).

19. The standard psychiatric reference for mental disorder classification is the *Diagnostic and Statistical Manual of Mental Disorders*. *See* AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS (5th ed. 2013) [hereinafter DSM-5]. The "Cautionary Statement" in part reads: "When DSM-5 categories, criteria, and textual descriptions are employed for forensic purposes, there is a risk that diagnostic information will be misused or misunderstood. These dangers arise because of the imperfect fit between the questions of ultimate concern to the law and the information contained in a clinical diagnosis." *Id.* at 25.

20. *See generally* David L. Faigman & Shayna Lewis, Anatomy of Scientific Evidence: Dissecting How Courts Reason from Group Data to Individual Decision Making (G2i) (May 27, 2014) (unpublished manuscript) (on file with authors).

an individual case to diagnostic labels that have been developed on the basis of group data and a collective clinical understanding. We suggest that outlining G2i-related issues in psychiatry will help judges, attorneys, and other participants in the legal system to better understand the interface between psychiatry and law, thus contributing to the better use of psychiatric evidence.

This Article employs the G2i lens to help bring into focus the point at which law and psychiatry intersect and, in the process, highlights a set of issues that will help judges, lawyers, and psychiatrists manage that intersection. Our intention is to provide illustrative examples, which neither represent a comprehensive "solution" to the topics addressed below nor identify all examples in which G2i issues are relevant to mental health law. Nonetheless, we do intend to provide a framework by which the role of psychiatric evidence in legal settings can better be understood. Through this Article, we strive to clarify the process of inference in psychiatric evidence, thus improving understanding of the "fit" of clinical concepts to legal questions. In some cases, examining the G2i issues will illuminate areas where the law might need to further define the fundamental questions that psychiatry is expected to answer. Ultimately, we propose that G2i inference is a useful lens through which the interactions of psychiatry and law can be better understood.

III. G2i in Law and Psychiatry

The basic insight associated with G2i inference—that is, that scientists focus on generalizing from group data and not individual cases—resonates in law and in psychiatry both independently and at the intersection of the two fields. G2i is inherent in the nature of applied science as well as in the procedural practices of the law. The law employs applied science in a wide variety of contexts—including product liability, criminal forensics, and medical malpractice. The courtroom setting is the focus of our inquiry here. Psychiatry, of course, is itself an applied science. Hence, as with all scientific disciplines, it begins by identifying certain general phenomena and then, in application, determining whether a particular case is an instance of that phenomenon. In this section, we describe how G2i inference operates in law and in psychiatry, respectively. In subsequent sections, we consider G2i when the two disciplines come together.

We begin with a brief explanation of terms. The notion of G2i captures the basic structure of applied science in that "G" refers to the group research data that might support a general hypothesis or statement about the world. Examples include, "a daily aspirin lowers the risk of heart attacks" or "the presence of a subdural hematoma in a dead infant is

consistent with abusive head trauma." As described by Faigman, Monahan, and Slobogin, in legal settings this level of science can be usefully referred to as "framework evidence."[21] Psychiatry, as an applied science, also begins with general statements about groups and presents this knowledge in diagnostic categories, sometimes catalogued in formal tomes—such as the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* ("*DSM-5*")—and sometimes not. Schizophrenia, bipolar disorder, and post-traumatic stress disorder ("PTSD") are examples of diagnostic categories included in the *DSM-5*.[22] But repressed memories, battered woman syndrome, and child abuse accommodation syndrome are examples of diagnostic categories that are not included in the *DSM-5* but are regularly used by both courts and mental health professionals.

In applied settings, however, the question is whether a particular case is an instance of the empirical framework or of a diagnostic category. This is the "i" of G2i. Hence, a psychiatrist would seek to determine, and a court might want to know, whether a particular party meets the criteria for schizophrenia or bipolar disorder, because different consequences might follow each diagnosis. Faigman, Monahan, and Slobogin referred to this determination in the law as "*diagnostic* evidence."[23] Similarly, psychiatrists describe their process of determining whether a particular case is an instance of a psychiatric category as a "diagnosis." Thus, in both law and in psychiatry the term "diagnostic" refers to the ability to determine whether a particular case is an instance of an empirical framework. But the scientific or legal relevance of making this determination depends on the context. The statement that John Smith has schizophrenia embodies a psychiatric diagnosis that is scientifically relevant for treatment (and possibly placement) purposes. In a courtroom in which the issue is whether John Smith suffers from a mental illness, the testimony that John Smith is a person with schizophrenia would be legally relevant as *diagnostic* evidence.

However, it should be emphasized at the outset that substantive law defines the content or meaning of relevant empirical framework evidence or *diagnostic* evidence, and these definitions may or may not align with the conventional diagnostic categories or diagnoses of psychiatry.

---

21. *See* Faigman, Monahan & Slobogin, *supra* note 3, at 424, 432.

22. *See generally* DSM-5, *supra* note 19.

23. *See* Faigman, Monahan & Slobogin, *supra* note 3, at 435. Although we adopt the terminology of Faigman et al. here, we recognize the potential for confusion between the category of individualized applications of group data that Faigman et al. refer to as "diagnostic" and the use of that term in psychiatry and other areas of medicine to indicate a person's particular diagnosis. To reduce the risk of confusion, in this Article we italicize *diagnostic* when using it to refer to the "i" in "G2i," but leave it unitalicized when using it in its medical sense.

There are many examples, some of which are explored in the sections below, but the classic example is the legal framework of "insanity." Insanity is a legal category that has no one-to-one corresponding category in psychiatry. Insanity is defined in many jurisdictions as a cognitive impairment at the time of the offense that rendered the defendant unaware of the wrongfulness of his conduct.[24] But in psychiatry there is no diagnostic category that corresponds fully to this definition of insanity. There are, however, many psychiatric categories or frameworks that are potentially relevant to the legal issue in dispute—i.e., was a particular defendant insane when he committed the crime? Schizophrenia, for instance, is one such category that raises the G2i issue of whether the defendant's condition warrants the legal classification of insanity. The problem is that the G2i of schizophrenia is not coterminous with the G2i of insanity. In this context, there is both a G2i of psychiatry (i.e., G=schizophrenia; i=defendant has schizophrenia) and a G2i of law (i.e., G=insanity; i=defendant is insane). Indeed, many jurisdictions permit psychiatric testimony regarding the defendant's mental state, which might involve whether the defendant suffers from schizophrenia, but expressly preclude testimony on the ultimate legal issue of the defendant's insanity.[25] Hence, a significant part of the challenge of developing a jurisprudence of psychiatric evidence lies in the need to translate between science and law.

Moreover, an additional translational challenge must be noted. Even when the courts and psychiatrists are using the same terms with the same meanings, the significance of each field's frameworks or diagnostic categories might be different in the two domains. For example, the diagnostic category of "intellectual disability" was developed for purposes of treatment and clinical decisionmaking regarding, for instance, appropriate placement in treatment and residential facilities of those so

24. *See, e.g.*, FLA. STAT. § 775.027 (2014) ("Insanity is established when . . . defendant did not know that what he or she was doing was wrong.").

25. *See, e.g.*, FED. R. EVID. 704(b) ("In a criminal case, an expert witness must not state an opinion about whether the defendant did or did not have a mental state or condition that constitutes an element of the crime charged or of a defense. Those matters are for the trier of fact alone."). However, some jurisdictions retain the previous federal rule allowing experts to testify to the ultimate legal questions. *See, e.g.*, N.Y. CRIM. PROC. LAW § 60.55(1) (McKinney 2014) ("When, in connection with the affirmative defense of lack of criminal responsibility by reason of mental disease or defect, a psychiatrist or licensed psychologist testifies at a trial concerning the defendant's mental condition at the time of the conduct charged to constitute a crime, he must be permitted to make a statement as to the nature of any examination of the defendant, the diagnosis of the mental condition of the defendant and his opinion as to the extent, if any, to which the capacity of the defendant to know or appreciate the nature and consequence of such conduct, or its wrongfulness, was impaired as a result of mental disease or defect at that time.").

diagnosed.[26] When the courts use the term "intellectual disability" to identify a class of offenders who are exempted from the death penalty, the term is employed to effectuate principles and values contained in the Eighth Amendment.[27] Whether the psychiatric diagnostic category of "intellectual disability" corresponds identically to the legal empirical framework contemplated by the Eighth Amendment is a translational issue that must be resolved by the courts.[28] Moreover, the different goals and consequences of applying diagnoses in clinical and legal settings might legitimately call for different levels of certainty before diagnostic conclusions are reached. For example, there may be varying risks of bias on the part of the evaluator and very different incentives for accuracy in reporting by the person being evaluated. Thus, in a legal context, a person may be incentivized to try to obtain a diagnosis that might help to mitigate his responsibility for his actions and thereby lessen the punishment. In contrast, a person in a clinical setting is generally less likely to want to receive a diagnosis of a serious mental illness.

A. *G2i in Law: Admissibility Standards*

Although the underlying empirical bases for framework and *diagnostic* expert testimony are likely to differ considerably, the same basic evidentiary rules of admissibility apply to both. As a generic matter,

26. *See, e.g.*, *Intellectual Disability*, ARC, http://www.thearc.org/learn-about/intellectual-disability/diagnosis (last visited Mar. 15, 2015).

27. *See* Atkins v. Virginia, 536 U.S. 304, 321 (2002) ("Construing and applying the Eighth Admendment in the light of our 'evolving standards of decency," we therefore conclude that such punishment is excessive and that the Constitution 'places substantive restriction on the State's power to take the life' of a mentally retarded offender.") (quoting Ford v. Wainwright 477 U.S. 399, 405 (1986)). *But cf.* Lois A. Weithorn, *Conceptual Hurdles to the Application of* Atkins v. Virginia, 59 HASTINGS L.J. 1203, 1206 (2008) (identifying the "conceptual challenges inherent in determining who is and who is not 'mentally retarded' for the purpose of applying the U.S. Supreme Court's 2002 decision in *Atkins v. Virginia*").

28. Indeed, the Supreme Court's recent decision in *Hall v. Florida* addressed exactly this issue and largely, though not entirely, adopted the consensus professional definition of intellectual disability. 134 S. Ct. 1986, 2000 (2014) ("In addition to the views of the States and the Court's precedent, this determination is informed by the views of medical experts. These views do not dictate the Court's decision, yet the Court does not disregard these informed assessments. . . . It is the Court's duty to interpret the Constitution, but it need not do so in isolation. The legal determination of intellectual disability is distinct from a medical diagnosis, but it is informed by the medical community's diagnostic framework.") (internal citation omitted). *But see id.* at 2006 (Alito, J., dissenting) ("[T]he Court binds Eighth Amendment law to definitions of intellectual disability that are promulgated for use in making a variety of decisions that are quite different from the decision whether the imposition of a death sentence in a particular case would serve a valid penological end. In a death-penalty case, intellectual functioning is important because of its correlation with the ability to understand the gravity of the crime and the purpose of the penalty, as well as the ability to resist a momentary impulse or the influence of others. By contrast, in determining eligibility for social services, adaptive functioning may be much more important.") (internal citation omitted).

most evidence codes require experts to be qualified and their testimony to be: (1) relevant to the legal issues in dispute; (2) helpful to the jury (or factfinder); and (3) supported by a reliable and valid foundation. We briefly consider each of these here, but return to them in detail in subsequent sections.

1. Common Requirements of "Qualifications," "Relevance," and Helpfulness

In most contexts, the qualifications requirement is straightforward and uncomplicated. Proffered experts must demonstrate sufficient knowledge of the subject of their testimony, which could have been gained through education, training, or experience. However, in the area of the psychological sciences, including psychiatry, this mandate is unevenly applied. Whereas some courts expect a high level of educational attainment, such as an M.D. or Ph.D.,[29] others permit lesser degrees with minimal training or experience to suffice.[30] At the least, however, courts should ensure that an expert's background is adequate to the demands of the testimony.[31]

If the expert is in fact qualified to testify, the court must next consider the content of his or her testimony. The first consideration is relevance, or what the *Daubert* Court referred to as "fit."[32] The fit criterion has two aspects in regard to scientific expert testimony: One can be described as the "legal fit," and the other as the "empirical fit." Legal fit refers to whether the expert's testimony concerns a fact that is at issue under the substantive law. For example, if the law does not allow a defense of "diminished capacity," expert testimony on that point regarding the defendant's intellectual disability would be deemed irrelevant.[33] Empirical fit, in contrast, refers to whether the research underlying the expert's testimony can help resolve the legal issues in dispute.[34] In

29. *See, e.g.*, Gebhardt v. Mentor Corp., 191 F.R.D. 180, 183 (D. Ariz. 1999) (excluding proffered expert from testifying on medical causation in part because the expert did not have a medical degree).

30. *See, e.g.*, McCoy v. Whirlpool Corp., 214 F.R.D. 646, 650 (D. Kan. 2003) ("As we know from the scarecrow in 'The Wizard of Oz,' a witness who is otherwise qualified is not disqualified for lack of credentials.").

31. *See* Schmerling v. Danek Medical, Inc., No. 96-cv-02749, 1999 WL 712591, at *9 (E.D. Pa. Sept. 10, 1999) (explaining that a physiatrist "that is, a specialist in physical or 'rehabilitation medicine,'" did not necessarily have the background that would permit him to opine on the alleged deficiencies of defendant's pedicle bone screw).

32. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 591 (1993).

33. In capital cases, however, at least at sentencing, intellectual disability is a relevant argument under the Eighth Amendment's guarantee against cruel and unusual punishment. *See* Atkins v. Virginia, 536 U.S. 304, 321 (2002) (citing Ford v. Wainwright, 477 U.S. 399, 405 (1986)).

34. *See, e.g.*, United States v. Birdsbill, 243 F. Supp. 2d 1128, 1131 (D. Mont. 2003)

social science terms, this is an issue of external validity. For example, research on predictions of future violence conducted entirely on populations of non-incarcerated adolescents might be found to not empirically fit a case involving the likelihood of violence of an adolescent housed in a secure facility.[35]

Expert testimony must not only be relevant, it must also add something to the factfinder's deliberations. It must "assist the trier of fact."[36] This helpfulness criterion is fairly modest in practice and is generally interpreted as requiring that the expert testimony add value beyond the factfinder's common experience or common sense.

2. Measuring Reliability and Validity

Most evidence codes require judges to operate as gatekeepers in regards to the scientific quality of expert scientific testimony. However the criteria that must be met to pass through the gate and how strictly those criteria are applied varies by jurisdiction and even by judge. Whatever expert rule courts might employ, it is intended as a measure of reliability and validity of the proffered evidence. The modern rule, as expressed in *Daubert*, explicitly embraces the gatekeeper metaphor and declares that judges have the responsibility to examine the methods and principles that underlie a proffered expert's testimony.[37] *Daubert* thus thrusts judges into the position of directly evaluating the reliability and validity of an expert's testimony. Since it was decided, *Daubert* has largely displaced the preexisting approach that was set forth in the venerable case of *Frye v. United States*,[38] which evaluated experts based on the general acceptance of their scientific testimony within their respective field.[39] *Frye* therefore operated as an indirect evaluation of reliability and validity, because it called upon courts to survey the applicable field for the answer. *Frye* and *Daubert* are often contrasted with one another, although in practice the principles embodied by these two

(excluding from evidence the results of a psychological test intended only to be used in treatment for pedophilia, not for the forensic purpose of detecting when "a person has sexually abused children").

35. *See, e.g.*, United States v. White Horse, 316 F.3d 769, 775 (8th Cir. 2003) (excluding from evidence a psychological test that measured sexual interest in children when the test excluded incest-only cases in its predictive equations "because incest offenders often act for reasons other than sexual interest").

36. Fed. R. Evid. 702; *Daubert*, 509 U.S. at 591.

37. *Daubert*, 509 U.S. at 592–93, 597.

38. 293 F. 1013 (D.C. Cir. 1923).

39. *Id.* at 1014 ("[T]he thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."). *See also* Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence:* Frye v. United States*, a Half-Century Later*, 80 Colum. L. Rev. 1197, 1204–05 (1980) ("*Frye* imposes a special burden—the technique must be *generally accepted by the relevant scientific community*.").

decisions are frequently applied in tandem. This section briefly outlines those principles. Later sections will apply them in considerable detail to psychiatric testimony.

i. *Frye* and Tests of "Acceptance"

Although there are many permutations of the *Frye* rule, the basic requirement can be simply stated: Under *Frye*, a trial court must determine whether proffered scientific evidence is generally accepted in the particular field to which it belongs.[40] In regard to the admissibility issues that arise under *Frye* concerning framework and *diagnostic* evidence, three issues in particular are worthy of note. Historically, judicial application of *Frye* created ambiguity around: (1) the proper level of generality at which *Frye* applies; (2) the correspondence between what is accepted in the field and what is sought to be testified to in court; and (3) the proper field for assessing acceptance.[41] We consider these each in turn.

Long before the issue of G2i was recognized, courts debated the proper level of analysis at which *Frye* applied. For example, if Magnetic Resonance Imaging ("MRI") is a "generally accepted" methodology, should courts also inquire into whether it is similarly accepted for the specific application to which it is being applied in the case at hand? MRI might be an excellent tool for diagnosing a brain tumor, but is its measurement of gray matter relevant to a claim of intellectual disability? And what about extensions of the basic technology, such as functional MRI ("fMRI")? With G2i, of course, this issue is multiplied, because *Frye* should perhaps require that judges ask whether the research underlying the framework evidence is generally accepted, as well as whether the science is sufficient to support the proffer of *diagnostic* evidence. Those courts still using the *Frye* test,[42] however, have not yet addressed this issue.

The second ambiguity of a *Frye* analysis—the correspondence between what is accepted in the field and what is sought to be proven in court—is particularly salient in the context of the law's use of psychiatric knowledge. Much of the corpus of knowledge in psychiatry is still not adequately researched and is not intended to serve forensic purposes. As noted before, the *DSM-5* expressly warns about this disjuncture.[43] This is not to say that psychiatric categories that are developed using

40. *Frye*, 293 F. at 1014.

41. *See generally* Gianelli, *supra* note 39.

42. While the federal circuit and a majority of state courts have adopted the *Daubert* standard, some states continue to favor the *Frye* standard. *See, e.g.*, Turner v. State, 746 So. 2d 355, 357 (Ala. 1998).

43. *See supra* note 19 and accompanying text.

sound scientific research might not have considerable relevance to questions arising under substantive law. But courts must be vigilant to ensure that the proper question is asked under *Frye*: Is this expert evidence generally accepted for the purpose for which it is being used in court? Scientific knowledge that is generally accepted for therapeutic purposes may or may not be generally accepted for forensic purposes.

The third ambiguity that surrounds *Frye*—the proper field in which the court should inquire regarding the general acceptance of proffered scientific evidence—plagues much expert evidence and is of particular concern in the field of behavioral sciences. Similar topics involving human behavior can be the subject of study of a myriad of fields. Consider, for example, the issue of implicit bias—a psychological concept that has been studied by psychiatrists, psychologists, neuroscientists, sociologists, and likely many clinicians with varied levels or kinds of training.[44] If these fields disagree about whether the phenomenon exists (i.e., the empirical framework), *Frye* provides no guidance regarding which field should be surveyed to assess general acceptance. This issue, of course, is multiplied manifold when the dispute concerns both framework and *diagnostic* evidence since the number of fields with possible opinions is multiplied as well.

ii. *Daubert* and Tests of "Validity"

A somewhat more challenging hurdle for much expert testimony is the requirement that it be based on a reliable and valid empirical foundation. The basic holding in *Daubert* was that under Rule 702 of the Federal Rules of Evidence, trial courts have the responsibility to determine whether, more likely than not, the expert's testimony is based on reliable methods and principles.[45] To assist courts in making this determination, the Court in *Daubert* and two subsequent cases (*General Electric Co. v. Joiner*[46] and *Kumho Tire Co. v. Carmichael*[47]) suggested criteria that might be considered. The Court instructed trial courts to inquire whether: (1) the theory or technique can be, and has been, tested; (2) the error rate is acceptable, and adequate standards exist to control the technique's operation; (3) the theory or technique has been peer reviewed and published; (4) there is "widespread acceptance" of the theory or technique;[48] and (5) the expert "employ[ed] in the courtroom the same

44. *See generally* Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945 (2006).

45. *Daubert*, 509 U.S. at 592 n.10 ("These matters should be established by a preponderance of proof.") (citing Bourjaily v. United States, 483 U.S. 171, 175–76 (1987)).

46. 522 U.S. 136 (1997).

47. 526 U.S. 137 (1999).

48. *Daubert*, 509 U.S. at 593–95.

level of intellectual rigor that characterizes the practice of an expert in the relevant field."[49] These factors, however, were not intended to be exclusive, and the *Daubert* Court emphasized that the inquiry under Rule 702 must be "a flexible one."[50]

The insight associated with G2i puts into question the allowable scope of proffered expert testimony. Whenever a scientific expert testifies, he or she must initially demonstrate the validity of the framework that is relevant to the legal dispute. This could involve a myriad of possible issues, from whether a particular substance can cause the kind of cancer from which the plaintiff suffers to the existence of the phenomenon of repressed memories. If the court finds that there is a sufficient empirical basis to support the proffered framework testimony, it should next consider whether the expert will be allowed to testify to whether the case at hand is likely an instance of that empirical framework. In the examples above, this would involve whether the substance at issue actually caused the plaintiff's cancer and whether the witness now recollects what was a previously repressed memory. Hence, all scientific expert testimony presents the two-part G2i question of whether there is a sufficient basis to support the proffered framework testimony and, if so, whether there is a further basis to support the proffered *diagnostic* testimony that this case is an instance of that framework.

It turns out that courts have not handled the issues arising with the G2i problem with any consistency or, indeed, have even recognized that they are of particular concern. In some areas of inquiry, courts routinely limit expert testimony to framework evidence alone. The example mentioned in Part I, expert testimony in eyewitness identification cases, is a particularly clear illustration of this approach. Courts that admit expert testimony regarding factors that can interfere with the accuracy of eyewitness identification almost invariably do not permit such experts to offer an opinion on whether a particular identification was accurate or not.[51]

In other contexts, however, courts require admissible expert testi-

---

49. *Kumho Tire Co.*, 526 U.S. at 152.

50. *Daubert*, 509 U.S. at 594.

51. *See, e.g.*, United States v. Hines, 55 F. Supp. 2d 62, 72 (D. Mass. 1999) ("The 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."); *see generally* Brian L. Cutler & Gary L. Wells, *Expert Testimony Regarding Eyewitness Identification*, *in* PSYCHOLOGICAL SCIENCE IN THE COURTROOM: CONSENSUS AND CONTROVERSY 100, 113 (Jennifer L. Skeem et al. eds., 2009) ("The main value of expert testimony is to educate the jury about eyewitness memory and the research findings. The state of the science . . . does not permit an assessment of the accuracy of an individual eyewitness. Accordingly, an opinion that a witness in a specific set of circumstances is unlikely to be accurate . . . is not a scientifically supported use of expert testimony.").

mony both on the existence of the framework and that the case at hand is an instance of that framework. For example, in medical causation cases, failure to provide valid proof of both "general causation" (i.e., framework evidence) and "specific causation" (i.e., *diagnostic* evidence) can result in a judgment against the proponent of the evidence.[52]

As discussed in greater detail in Part III, psychiatry presents all of the same G2i issues in the law as do the physical, biological, and social sciences, since psychiatry is something of an amalgam of all of these. However, because psychiatry is relevant to a person's mental state, which is often a relevant element under the law, it also confronts additional admissibility issues beyond those that other kinds of scientific expert testimony confront. But before reaching the issue of how psychiatric evidence fits into legal decisionmaking, it behooves us to consider how G2i manifests specifically within psychiatry as a research and clinical discipline.

B. *G2i in Psychiatry*

The field of psychiatry, like all applied sciences, is an exercise in identifying general phenomena for a variety of purposes, including but not limited to classification, association, and causation. For example, research on schizophrenia entails defining the phenomenon of interest, examining the factors associated with its existence (i.e., onset, severity, and so forth), and then—ideally—identifying treatments that might ameliorate its effects. All such research on schizophrenia as a phenomenon considers it at the population level, although the data are aggregated from individual cases. Yet the primary use of this knowledge about schizophrenia in the field of psychiatry is clinical. Thus, as a routine matter of professional practice, psychiatrists conduct G2i inference. This section examines this professional practice, while the following sections will consider this practice as it fits (or does not fit) the law's requirements.

1. *Diagnostic* Frameworks

All standard psychiatric diagnoses rely on a set of diagnostic criteria delineated in the *Diagnostic and Statistical Manual of Mental Disorders*, now in its fifth edition, published by the American Psychiatric Association.[53] For each mental disorder, the *DSM-5* describes an organized list of criteria composed of characteristic symptoms (i.e., subjec-

52. *See, e.g.*, *In re* Aredia & Zometa Prod. Liab. Litig., 483 F. App'x 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.").

53. DSM-5, *supra* note 19, at xli.

tive problems reported by the person, such as sadness or anxiety) and signs (i.e., observations made by the clinician, such as the patient's appearance or cognitive function) that characterize the syndrome in question.[54] In order to receive a particular diagnosis, in most cases the person must meet a certain number of those criteria (e.g., two or more).[55] Most diagnoses also specify duration and frequency requirements (e.g., "at least six months"),[56] and most disorders require the evaluator to make a judgment about the actual impact of the disorder on the subject's functioning (e.g., "clinically significant distress or impairment in social, occupational, or other important areas of functioning").[57]

---

54. For example, the diagnosis of schizophrenia requires that the person meet the following criteria:

> A. Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated). At least one of these must be (1), (2), or (3):
>
> 1. Delusions.
> 2. Hallucinations.
> 3. Disorganized speech (e.g., frequent derailment or incoherence).
> 4. Grossly disorganized or catatonic behavior.
> 5. Negative symptoms (i.e., diminished emotional expression or avolition).
>
> B. For a significant portion of the time since the onset of the disturbance, level of functioning in one or more major areas, such as work, interpersonal relations, or self-care, is markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, there is failure to achieve expected level of interpersonal, academic, or occupational functioning).
>
> C. Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms (or less if successfully treated) that meet Criterion A (i.e., active-phase symptoms) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifested by only negative symptoms or by two or more symptoms listed in Criterion A present in an attenuated form (e.g., odd beliefs, unusual perceptual experiences).
>
> D. Schizoaffective disorder and depressive or bipolar disorder with psychotic features have been ruled out because either 1) no major depressive or manic episodes have occurred concurrently with the active-phase symptoms, or 2) if mood episodes have occurred during active-phase symptoms, they have been present for a minority of the total duration of the active and residual periods of the illness.
>
> E. The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition.
>
> F. If there is a history of autism spectrum disorder or a communication disorder of childhood onset, the additional diagnosis of schizophrenia is made only if prominent delusions or hallucinations, in addition to the other required symptoms of schizophrenia, are also present for at least 1 month (or less if successfully treated).

Id. at 99. Specifiers can also be added to indicate the course of the condition (e.g., acute, partial remission, full remission) and the severity, as well as the presence of catatonic symptoms. *Id.* at 99–100.

55. *Id.*

56. *See, e.g.*, *id.*

57. However, it is worth noting that from *DSM-IV-TR* (the fourth edition with revised text) to the current *DSM-5*, fewer disorders have this "clinical significance" criterion. *See* Paul S. Appelbaum, *Reference Guide on Mental Health Evidence*, *in* REFERENCE MANUAL ON SCIENTIFIC

In addition to standard (i.e., *DSM*-recognized) diagnoses, psychiatrists may testify to the existence of syndromes that may have legal significance but have not (or at least not yet) achieved general recognition by the field. These categories include purported behavioral syndromes such as "battered woman syndrome"[58] or "child sexual abuse accommodation syndrome,"[59] which are not typically the focus of psychiatric attention in the clinical setting and are thus unlikely to ever appear in the *DSM*. Legally significant categories of syndromes also often include proposed clinical syndromes for which research support is still being generated, which means that they lack sufficient evidence to be incorporated into the *DSM*. An example is hypersexual disorder (i.e., "sex addiction"), a concept with some degree of "face validity," given that some clinicians routinely provide treatment for it,[60] but which was rejected for inclusion in the *DSM-5* because of insufficient empirical support.[61] Diagnostic criteria have been proposed for most of these additional syndromes, but there may be conflicting diagnostic schemas and dispute over which is authoritative, or whether the syndrome is a disorder that can be cleanly distinguished from other disorders.

Of course the actual process of making a diagnosis requires more than simply counting the number of diagnostic criteria met. Clinical assessment is a process of gathering and synthesizing data that takes place in the context of a complete examination, including a thorough exploration of the person's personal history, social and occupational experiences, family history, medical status, and a clinician's objective findings, such as the mental status examination (a structured assessment of the person's mental state) and physical data (e.g., a physical examination, laboratory tests). Sometimes structured diagnostic interviews—i.e., systematized sets of questions that often include a guide for interpreting the results—are used to ensure that important details are not missed and to standardize the assessment process. Ultimately, clinicians must make

---

EVIDENCE (3d ed. 2011) (indicating that under *DSM-IV-TR*, "for a diagnosis of any disorder to be made per the *DSM*, the symptoms must be deemed to 'cause clinically significant distress or impairment in . . . important areas of functioning.'"); *see also* Am. Psychiatric Ass'n, *Psychiatric Evaluation of Adults, Second Edition*, 163 AM. J. PSYCHIATRY 3 (2006).

58. *See* Regina A. Schuller & Neil Vidmar, *Battered Woman Syndrome Evidence in the Courtroom: A Review of the Literature*, 16 L. & HUM. BEHAV. 273, 277 (1992).

59. *See generally* Kenneth J. Weiss & Julia Curcio Alexander, *Sex, Lies, and Statistics: Inferences from the Child Sexual Abuse Accommodation Syndrome*, 41 J. AM. ACAD. PSYCHIATRY & L. 412 (2013).

60. Robert Weiss, *No Sex Please, We're Psychiatrists*, PSYCH CENT., http://blogs.psychcentral.com/sex/2012/12/hypersexual-disorders-dsm/ (last visited Mar. 28, 2015) (noting that many "Masters and PhD level clinicians . . . are treating sexually addicted clients on a regular basis").

61. *See* Paul S. Appelbaum, *Commentary: DSM-5 and Forensic Psychiatry*, 42 J. AM. ACAD. PSYCHIATRY & L. 136, 137 (2014).

assessments as to the "clinical significance" of the observed symptoms and their effect on functioning.

The field of psychiatric nosology (the classification of diseases) is currently at the center of an active debate about the definition of mental disorder—especially in light of the recent revision of the *DSM*—and disagreement over certain underlying conceptual issues that are particularly relevant to this discussion.[62] For example, the concept of *validity* is the extent to which a measurement corresponds to the "real world." In psychiatry, validity implies that a given diagnostic label accurately defines a discrete syndrome that reflects a particular mental state.[63] With regard to validity, ontology (do disorders exist?) can be framed as a separate question from epistemology (how can we know anything about them if they do exist?).[64] Many psychiatric researchers and practitioners accept the first point—that mental disorders do exist and are real phenomena in the world—while having less confidence about the second point—i.e., there is recognition that psychiatrists cannot always accurately perceive and describe mental disorders as they exist in the world.[65] This difficulty in understanding and characterizing the fundamental nature of disorders is reflected in different conceptions of validity that demonstrate areas of possible disjunction between measurement and reality: content validity (does the diagnosis measure all the generally accepted characteristics of the disorder construct?), concurrent validity (does it correlate with other findings associated with the disorder, such as laboratory findings or brain imaging?), discriminant validity (does it delimit this disorder from other disorders and from the absence of disorder?), and predictive validity (does it add to the ability to anticipate outcomes?).[66]

Diagnostic criteria are assessed more concretely through measures of *accuracy,* comprising the statistical measures of *sensitivity* and *specificity*. These characteristics can be described in any diagnostic instrument or test, and they correspond respectively to a test's ability to "rule in" people who actually have the condition in question and to "rule out"

---

62. James Phillips et al., *The Six Most Essential Questions in Psychiatric Diagnosis: A Pluralogue, Part 1: Conceptual and Definitional Issues in Psychiatric Diagnosis*, 7 PHIL. ETHICS & HUMAN. MED., Jan. 2012, at 1, 3, 17, *available at* http://www.peh-med.com/content/7/1/3 [hereinafter *Six Most Essential Questions*].

63. Appelbaum, *supra* note 57, at 839 (defining validity as "the extent to which the diagnosis corresponds to the person's actual mental state").

64. Claire Pouncey, Commentary, *Mental Disorders, Like Diseases, Are Constructs. So What?*, *in Six Most Essential Questions*, *supra* note 62, at 6.

65. *See generally, e.g.*, Phillips et al., *Six Most Essential Questions*, *supra* note 62.

66. *See* Eduard Vieta & Mary L. Phillips, *Deconstructing Bipolar Disorder: A Critical Review of Its Diagnostic Validity and a Proposal for* DSM-V *and* ICD-11, 33 SCHIZOPHRENIA BULL. 886, 886 (2007).

people who do not. In other words, sensitivity tells us the following: Of those persons who have the condition, what proportion will have a positive test result?[67] Similarly, specificity tells us what proportion of those persons who do not have the condition will have a negative test result?[68] Consider the example of a group of surgeons assessing patients with complaints of abdominal pain, where the surgeons are attempting to determine in which cases the pain is due to appendicitis. When the surgeons apply a standard set of criteria, some will likely miss an inflamed appendix that is actually there, and some may incorrectly diagnose a case of appendicitis that does not exist. Likewise, psychiatric diagnostic criteria will misidentify some "false positives" and "false negatives." There is also a tradeoff between these two characteristics: More inclusive tests increase sensitivity at the expense of specificity, and more restrictive tests do the opposite.[69] Judges who are acting as the "gatekeepers" envisioned by the *Daubert* Court will need to consider these issues of validity and accuracy (i.e., error rate) in assessing the admissibility of proffered psychiatric testimony.

Another set of empirical data about *DSM* diagnoses relates to *reliability*. Diagnostic reliability is a measure of repeatability and consensus. More specifically, interrater reliability is the probability of two clinicians examining the same person and deriving the same diagnosis, while test-retest reliability is a related measure that measures the stability of the diagnostic label over time.[70] Both concepts relate to replicability. Statistically, the reliability of *DSM* diagnoses is commensurate with the diagnostic reliability of other medical procedures, such as evaluating angiograms or reading x-rays, which is to say, better than chance but far from perfect.[71]

Reliability is often described using a statistical measure called kappa (κ), which corrects for chance agreement, whereby $\kappa = 0$ means there is no interrater reliability and $\kappa = 1$ indicates perfect interrater reliability; $\kappa = 0.5$ is approximately midway between chance expectation and 100% agreement.[72] Most medical reliability studies, including psychiatry and other branches of medicine, have generally found interrater

67. *See* Helena Chmura Kraemer et al., Commentary, *DSM-5: How Reliable Is Reliable Enough?*, 169 AM. J. PSYCHIATRY 13, 13 (2012).

68. *See id.*

69. *Id.* ("Any effort to improve the sensitivity of DSM-IV criteria will result in higher prevalence rates, and any effort to improve the specificity of DSM-IV criteria will result in lower prevalence rates.").

70. *Id.* at 13–14.

71. *See id.* at 14.

72. *See* Mary L. McHugh, *Interrater Reliability: The Kappa Statistic*, 22 BIOCHEMIA MEDICA 276 (2012), *available at* http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/.

kappa values between 0.4 and 0.6.[73] The most recent field trials conducted as part of the *DSM-5* revision process examined several diagnostic categories, of which five diagnoses were in the "very good" range ($\kappa = 0.60$ to 0.79), nine were in the "good" range ($\kappa = 0.40$ to 0.59), six were in the "questionable" range ($\kappa = 0.20$ to 0.39), and three were in the "unacceptable" range ($\kappa < 0.20$).[74] Indeed, as early as the late 1970s, field trials for *DSM-III* showed relatively strong interrater agreement.[75] Therefore, although some diagnostic categories appear to be less reliable than others, the majority of psychiatric diagnoses demonstrate degrees of reliability that are generally useful and not grossly out of sync with the rest of clinical science.[76] However, as discussed in greater detail below, the degree of reliability of the particular diagnostic category in question and of the diagnostic process used by the psychiatric expert represents another set of considerations for judges assessing admissibility.

The process of psychiatric diagnosis epitomizes the G2i process, as the group data on which diagnostic categories are based are applied to individual patients, or in the forensic context to defendants or plaintiffs. Considerations of both validity and reliability are relevant here. Diagnoses are developed on the basis of data indicating that their characteristic symptom clusters "hang together" (i.e., are observed together) in a number of ways, and thus are likely to represent valid constructs, and that the relevant symptoms can be identified in a sufficiently reliable fashion. The information that may be sought in the process of validating a diagnosis falls into three broad categories: antecedent validators (e.g., clustering of the diagnosis in families and shared environmental risk factors); concurrent validators (e.g., consistency across affected persons in the cognitive, emotional, temperamental, and personality correlates of the disorder); and predictive validators (e.g., diagnostic stability over

73. *See* Kraemer et al., *supra* note 67, at 14.

74. Of note, these data derive from application by clinicians who were not given training beforehand (i.e., they were often applying novel diagnostic formulations with which they lacked prior experience), so these trials were probably a conservative overestimate of the likely number of discrepancies between clinicians. *See* Darrel A. Regier et al., *DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses*, 170 AM. J. PSYCHIATRY 59, 59, 62 (2013).

75. CHRISTOPHER SLOBOGIN ET AL., LAW AND THE MENTAL HEALTH SYSTEM: CIVIL AND CRIMINAL ASPECTS 443–44 (5th ed. 2009) (indicating that "field tests of the new diagnostic scheme introduced by DSM-III . . . showed that for some major diagnostic categories the interrater agreement was quite high . . . .").

76. *See* Kraemer et al., *supra* note 67, at 14 ("It is unrealistic to expect that the quality of psychiatric diagnoses can be much greater than that of diagnoses in other areas of medicine, where diagnoses are largely based on evidence that can be directly observed. Psychiatric diagnoses continue to be based on inferences derived from patient self-reports or observations of patient behavior. Nevertheless, we propose that the standard of evaluation of the test-retest reliability of DSM-5 be consistent with what is known about the reliability of diagnoses in other areas of medicine.").

time, a predictable course of illness, and common responses to treatments).[77] Thus, making a diagnosis of an individual presumes that the aggregate data by means of which the diagnostic category was formulated applies equally well to this person, and that the level of reliability established for making this diagnosis in general applies to the process by which this specific diagnosis was made.

Yet once again, what is true in the aggregate for members of a large group may not be equally valid for a single person. In general, for example, most people who experience two of the following symptoms, including one of the first three symptoms over a period of at least six months, which cannot be accounted for by substance abuse or a general medical condition, are highly likely to have schizophrenia: delusions; hallucinations; disorganized speech; grossly abnormal psychomotor behavior, such as catatonia; and negative symptoms, i.e., restricted affect or avolition/asociality—along with impairment of social or occupational functioning.[78] Yet, it is not difficult for any experienced psychiatrist to recall patients who received a diagnosis of schizophrenia at some point in time but, in retrospect, were found not to be suffering from the disorder. What was thought to be schizophrenia may have turned out to be another psychiatric condition, or the symptoms may have resolved and, without further explanation, the person no longer appears to be experiencing a disorder at all. Even very strong data on the validators of a set of diagnostic criteria will, at best, only produce heavy odds that a diagnosis applies to an individual, but certainty is elusive. Moreover, although data exist suggesting that the relevant symptoms can be ascertained reliably, particular characteristics of a given evaluation (e.g., uncooperativeness of the evaluee) may reduce that reliability. Hence, psychiatric diagnosis is essentially a probabilistic statement about a person's likely diagnosis, rather than an absolute statement of fact.[79]

It is worth emphasizing that what is true for psychiatric diagnoses is true as well for all of medicine, and indeed for almost any situation in which group data are being applied to individuals—that is the essence of the G2i dilemma.

---

77. *See* Kenneth Kendler et al., *Guidelines for Making Changes to DSM-V*, AM. PSYCHIATRIC Ass'N, http://www.dsm5.org/about/Documents/DSM5%20Guidelines%20for%20Change.pdf (last revised Oct. 21, 2009).

78. DSM-5, *supra* note 19, at 99; *see also supra* note 54.

79. *See* John Ruscio, *Diagnoses and the Behaviors They Denote: A Critical Evaluation of the Labeling Theory of Mental Illness*, 3 SCI. R. MENTAL HEALTH PRAC., no. 1, Summer 2009, *available at* http://www.srmhp.org/0301/labels.html ("In fact, even highly reliable, valid, and useful [diagnoses] will only imperfectly describe the past and only probabilistically predict the future.").

2. Functional Assessment

The assessment of functional impairment is its own enterprise separate from psychiatric diagnosis. Functional capacities can include a broad range of cognitive and behavioral abilities, including perception, insight, or control of behavior.[80] Mental disorders can influence these capacities in a variety of ways, and a disorder diagnosis is useful to functional assessment insofar as it can suggest areas to explore in greater depth.[81] However, a diagnosis is usually insufficient to draw conclusions about a given functional impairment because "a broad range of functional impairments can be associated with almost any mental disorder."[82]

Functional impairment in the past, present, or future is usually the key issue in legal determinations involving psychiatric testimony. Was the defendant substantially unable to control his behavior at the time of the crime? Did the deceased testator have the capacity to alter her will? Is this person so impaired as to be unable to perform the basic functions of her job?[83] Does this father have the capacity to parent his child? Will this sexual offender recidivate?[84] Can treatment with medication restore this defendant's competence to proceed? Group data, such as those associated with diagnostic categories, can suggest areas in which impairment may have been present, may be present today, or may appear in the future, but given individual variability, these group distinctions will not be conclusive in an individual case. It is thus important to consider the nature of functional assessment and the potentially overlooked G2i issues inherent in this process.

Although the most basic examples of functional assessment can appear to be simply descriptive, thus avoiding G2i issues to some degree, functional assessment frequently requires judgment based in part on inference from group data. For example, reporting that a person cannot concentrate enough to complete simple tasks may require only straightforward observation and description of behavior. But often a functional assessment requires an evaluator to form an opinion regarding function in the past or future (e.g., capacity to make a decision or ability

80. *See* Appelbaum, *supra* note 57, at 841–42.

81. *Id.* at 843.

82. *Id.*

83. *See, e.g.*, Cravens v. Blue Cross & Blue Shield of Kan. City, 214 F.3d 1011, 1016 (8th Cir. 2000) ("The determination of qualification [under the American with Disabilities Act] involves [inquiry into] . . . (2) whether the individual can perform the essential job functions, with or without reasonable accommodation.").

84. *See, e.g.*, *In re* Anderson, 730 N.W.2d 570, 580 (2007) (Expert testifying that "[defendant's] disorder is logically consistent with there being a *causative connection* between a requisite mental condition and subsequent recidivist sexually predatory conduct.").

to parent a child). Inferential issues manifest in the space between data collection and the formation of an opinion regarding function. For example, a functional assessment might rely in part on a person's report of impairment, but it will likely also require gathering collateral information from other sources, considering information about performance in other areas of function (e.g., school, work, or military records), and ultimately forming an opinion about the overall congruence of clinical symptoms and alleged impairment.[85] Furthermore, to be legally relevant, the impairment typically must be linked to a disorder (e.g., disability evaluation of an employee who complains that anxiety disorder limits her ability to function at work) or the consequences of an event (e.g., litigation over a claim that an auto accident caused post-traumatic stress disorder, which in turn led to disabling anxiety).[86] This will require application of data about groups of people in similar circumstances to conclusions about the person in question (e.g., how frequently are people with major depression unable to work, or how often do automobile accidents result in post-traumatic stress disorder?).

There are several types of structured instruments for functional assessments, although many evaluations are conducted simply on the basis of a systematic clinical evaluation. At the most basic level, these instruments help to gather relevant information by organizing data collection, such as by providing a checklist of questions to ask or issues to consider.[87] It is then up to the examiner to make the necessary inferences connecting this information to an opinion regarding function (though, as noted above, even in this case the examiner would often be required to link this functional impairment to a diagnosis of a specific disorder). Some instruments additionally quantify the likelihood that a person will manifest a particular behavior or impairment by generating numerical scores that can be compared to population norms for enhanced predictive accuracy.[88] This type of instrument can raise substantial questions about the applicability of the group norms to the person being

85. *See* GARY B. MELTON ET AL., PSYCHOLOGICAL EVALUATIONS FOR THE COURTS 43–67 (3d ed. 2007) (suggesting that "because the legal system is interested in the just adjudication of disputes, accuracy is the primary goal" in determining the existance of an alleged impairment, and therefore "information should be sought not only from the examinee but from *all* relevant sources, including sources that the examinee may not know about or would rather not have consulted").

86. *See* Appelbaum, *supra* note 57.

87. *See, e.g.*, Kevin S. Douglas et al., *Historical-Clinical-Risk Management-20, Version 3 (HCR-20^{V3}): Development and Overview*, 13 INT'L J. FORENSIC MENTAL HEALTH 93, 98 tbl.1 (2014), *available at* http://www.tandfonline.com/doi/pdf/10.1080/14999013.2014.906519 (providing a list of violence risk factors).

88. *See, e.g.*, Marnie E. Rice & Grant T. Harris, *Violence Risk Appraisal Guide (VRAG)*, *in* 2 ENCYCLOPEDIA OF PSYCHOLOGY & LAW 847–48 (Brian L. Cutler ed., 2008) (describing the VRAG as "a 12-item actuarial instrument that assesses the risk of violent recidivism among men apprehended for criminal violence").

assessed.[89] Finally, some structured assessments involve direct testing of a person's capacity to perform particular tasks, for example, by directly observing skills relevant to driving,[90] money management,[91] or reasoning about medical decisions.[92] This last type of assessment mitigates G2i issues to some extent, but only insofar as the assessment questions address present functioning. When the relevant legal question relates to past or future functioning, inference from group performance typically will be involved.

Sometimes the legally relevant question will demand inferences about causality. For example, if a plaintiff claims impaired concentration as a manifestation of anxiety secondary to post-traumatic stress disorder caused by an automobile accident, psychiatric experts may be called upon not merely to identify the level of impairment, but also to link it to a specific cause.[93] It is significantly more challenging to consider the degree to which an observed impairment can be attributed to a specific cause, in part because answering questions of etiology is not part of the usual clinical enterprise in psychiatry. Mental health practice is ordinarily concerned with the observation and diagnosis of disorder, not the determination of causes of dysfunction. Differential etiology is a legal—not clinical—concept that describes a more direct determination of cause, in which the putative cause is first ruled in, then other causes are ruled out.[94] Mental disorder diagnoses may *suggest* etiology, at least in the sense that a mental disorder often correlates with certain etiologic factors (e.g., depression following loss of a loved one). But the mere presence of a diagnosis is by no means sufficient to make strong claims about its cause, and mental disorder diagnoses should not be taken as equivalent to claims of etiology.

---

89. *See* discussion *infra* Part IV.B.3.i.a.

90. *See generally* Maria T. Schultheis et al., *The Neurocognitive Driving Test: Applying Technology to the Assessment of Driving Ability Following Brain Injury*, 48 REHABILITATION PSYCHOL. 275 (2003).

91. *See* Daniel C. Marson et al., *Assessing Financial Capacity in Patients with Alzheimer Disease: A Conceptual Model and Prototype Instrument*, 57 ARCHIVES NEUROLOGY 877, 879 (2000).

92. *See, e.g.*, THOMAS GRISSO & PAUL S. APPELBAUM, MACARTHUR COMPETENCE ASSESSMENT TOOL FOR TREATMENT (MACCAT-T) 1 (1998) (describing the MacCAT-T as a tool that "offers physicians and other health professionals practical guidance in their assessments of patients' decisionmaking capacities in the context of informed consent to treatment").

93. For an example of a case where physchiatric experts were called to determine causality, see *State Office of Risk Management v. Larkins*, 258 S.W.3d 686, 691 (Tex. App. 2008) ("[W]e hold that expert testimony was required to establish that Larkins's depression, anxiety, and post-traumatic stress disorder were causally related to her head injury and therefore compensable . . . .").

94. *See* David L. Faigman & Claire Lesikar, *Organized Common Sense: Some Lessons from Judge Jack Weinstein's Uncommonly Sensible Approach to Expert Evidence*, 64 DEPAUL L. REV. 421 (2015).

Even when a conclusion about a diagnosis is not formally required in a given legal setting (e.g., adjudications of claims of incompetence or emotional harm), a diagnosis may play a role in bolstering the strength of an expert's testimony about functional impairment or emotional distress. That seems to be the view of many courts, which require experts to provide diagnoses before testifying regarding functional abilities, and of many attorneys, who ask experts to offer diagnostic formulations.[95] These practices appear to be based on the assumption that the identification of symptoms or functional abilities may be less reliable when such observations are not linked to a psychiatric diagnosis.[96]

Finally, translation issues may arise with regard to functional assessment. A process of assessment initially developed for a clinical context will need to be translated for a legal context, and G2i issues may be manifest in this process of translation, depending on the nature of the assessment. Psychiatry and related disciplines define certain functions and capacities for purposes of measurement, e.g., "adolescent development," "competence," or "intelligence." Because the core functional concepts are defined, initially, for psychiatric and not legal purposes, some degree of inference will be required in that process of translation.[97] The *DSM*'s "Cautionary Statement," warning that psychiatric concepts that are appropriate for clinical or research purposes may not map directly onto legally relevant categories,[98] is just as material for functional assessment as it is for diagnostic categorization. Given the heterogeneous nature of functional assessment, which can vary according to both the individual's clinical characteristics and the function in question, that translation will need to be assessed on a case-by-case basis. G2i analysis of the foundations of the aggregate data on which conclusions from the assessment rely are likely to be helpful in this regard.

IV. Psychiatry in the Courtroom: A User's Guide

Given the complexity surrounding the issue of the admissibility of psychiatric evidence, it is not surprising that the current state of affairs in both state and federal courts is something of a muddle. Using G2i guideposts, this section seeks to offer some guidance out of the muddle. It begins with a brief introduction of current approaches to psychiatric expert testimony. It then provides an extensive user's guide to such evi-

95. *See* Appelbaum, *supra* note 57, at 820 (citing 2 E. Allen Farnsworth, Contracts §§ 4.6–4.8, at 228–34 (4th ed. 2004)).

96. Although this is a plausible proposition, we are unaware of any empirical data addressing the question. This may be a useful area for future research.

97. Morse, *supra* note 11, at 889.

98. DSM-5, *supra* note 19, at 25.

dence, suggesting, in particular, best practices for both framework and *diagnostic* psychiatric testimony.

A. *Current Approaches to Admissibility*[99]

The definition of an expert and the evidentiary standards applied to the scientific testimony of mental health professionals have been subject to extensive discussion.[100] Admissibility of testimony will depend in part on how evidence about a diagnosis is connected to the legal issue in contention, including criminal responsibility (mental state during past actions), propensity (predictions about behavior and response to treatment), and competence (abilities to function).[101] However, here we focus primarily on admissibility concerns that arise from the nature of the diagnostic method and approaches to functional assessment.

Critics have argued that psychiatric expert testimony is unreliable, as there may be considerable variation across examiners and no clear benchmark for determining accuracy, and therefore it should be considered insufficient to meet evidentiary standards like *Frye* and *Daubert.*[102] It has also been suggested that diagnostic categories are sufficiently heterogeneous that they convey little more information than would a simple description of the symptoms on which they are based.[103] Concerns regarding the validity of a number of psychiatric diagnostic categories have also been expressed. In practice, though, forensic mental health professionals rarely have their testimony excluded or significantly limited on such grounds.[104] Testimony that relies on diagnoses found in the

99. To focus this discussion, we restrict it to a limited number of legal contexts. Even so, the vast array of possible applications of mental health evidence makes an exhaustive description of all G2i considerations relevant to these topics impossible. We also recognize that diagnosis and assessment cannot be entirely cordoned off from other questions in mental health law. Legal arguments incorporating a mental disorder predicate often require diagnosis to be connected to the relevant legal issues, such as culpability or the prediction of behavior, so ultimately the issues of diagnosis and functional assessment cannot be considered in isolation.

100. *See generally* MELTON ET AL., *supra* note 85, at 3–24.

101. *See* SLOBOGIN ET AL., *supra* note 75, at 424.

102. *See* Michael H. Gottesman, *Admissibility of Expert Testimony After* Daubert*: The "Prestige" Factor*, 43 EMORY L.J. 867, 875 (1994) ("Are psychiatrists' assessments of the mental capacity of a defendant at the time of the crime 'testable' or 'falsifiable' or 'refutable'? Plainly not. Can we determine the 'error rate' of psychiatric opinion, or utilize standards to control the technique's operation? Again, plainly not.").

103. *See* Morse, *supra* note 11, at 889 ("[T]here is enormous heterogeneity within each disorder category. That is, people who technically meet the criteria for the diagnosis may have quite different presentations.").

104. *See* CHRISTOPHER SLOBOGIN, PROVING THE UNPROVABLE 27 (2007). There may be some variation in the level of scrutiny to which psychiatric evidence is subjected based on the legal context. Courts in child custody proceedings, for example, tend to be quite permissive about the introduction of expert testimony. *See* Elizabeth S. Scott & Robert E. Emery, *Gender Politics and Child Custody: The Puzzling Persistence of the Best-Interests Standard*, 77 L. & CONTEMP. PROBS. 69, 92 (2014) ("[T]he rules that generally restrict the admissibility of scientific evidence in

DSM is particularly well accepted.[105] Furthermore, the uncommon exclusions of psychiatric testimony that have occurred after *Daubert* appear to be on the grounds of relevance, not reliability.[106]

The fundamental justification for why *DSM*-backed psychiatric evidence is generally exempt from these screening analyses is subject to competing interpretations, and in practice courts are likely not consistent in their reasons for taking a relatively lax approach to such evidence. One possibility is that behavioral testimony is not perceived as being based on "hard science," and therefore it does not need to be scrutinized to the same degree as evidence drawn from the physical sciences.[107] Similarly, though in a somewhat more nuanced line of reasoning, it has been suggested that an understanding of mental health testimony "is accessible to the jury, and not dependent on familiarity with highly technical or obscure scientific theories," and therefore that "the expert's qualifications, and the logical bases of his opinions and conclusions can be effectively challenged by cross-examination and rebuttal evidence."[108] Finally, there is a long tradition of finding mental health testimony admissible, and courts may simply be respecting precedent.

In some jurisdictions, psychiatric and psychological expert evidence is sometimes explicitly divided into two basic categories: opinions based on research and statistical tests, and opinions based on professional experience. The former are treated with caution and closer evi-

---

legal proceedings are often not applied to custody proceedings, and judges tend to be uncritical in assessing the quality of the opinions of court-appointed experts."). Testimony in criminal cases, however, may be subject to greater scrutiny, especially when it involves syndromes not reflected in the *DSM*.

105. For example, "as of 2001 . . . 'no reported *Daubert* challenges to retrospective psychiatric assessments of criminal responsibility' that rel[ied] on the DSM" could be found. SLOBOGIN, *supra* note 104, at 27 (citation omitted).

106. *See, e.g.*, Henry F. Fradella et al., *The Impact of* Daubert *on the Admissibility of Behavioral Science Testimony*, 30 PEPP. L. REV. 403, 421, 434, 437 (2003).

107. This approach of distinguishing scientific expert testimony from non-scientific expert testimony in regard to the court's gatekeeping responsibilities was rejected by the Court in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 147 (1999) ("The initial question before us is whether this basic gatekeeping obligation applies only to 'scientific' testimony or to all expert testimony. We, like the parties, believe that it applies to all expert testimony."). Some states, however, maintain this distinction and apply a more lenient test of admissibility for expert testimony that does not claim a scientific basis. *See, e.g.*, Marron v. Stromstad, 123 P.3d 992, 1004 (Alaska 2005) ("But we have never adopted *Kumho Tire's* extension of *Daubert* to all expert testimony . . . . [W]e limit our application of *Daubert* to expert testimony based on scientific theory, as opposed to testimony based upon the expert's personal experience."). *But see* MELTON ET AL., *supra* note 85, at 3 ("The public's antipathy toward clinical opinion appears to stem from the belief that much 'expert' testimony is based on 'junk science' . . . .").

108. State v. Borrelli, 629 A.2d 1105, 1111 (Conn. 1993) (finding expert's battered woman syndrome testimony admissible); *see also* Maxine D. Goodman, *A Hedgehog on the Witness Stand—What's the Big Idea?: The Challenges of Using* Daubert *to Assess Social Science and Nonscientific Testimony*, 59 AM. U. L. REV. 635, 654, 654 n.99 (2010).

dentiary review, while the latter are simply considered "inductive" and do not receive such close examination. In the California case of *People v. Miller*, for example, a state appellate court found that clinical testimony predicting future violence under the state's Sexually Violent Predator Act was not scientific evidence and therefore not subject to California's version of the *Frye* test, even though the testimony made use of statistical data.[109] The defendant argued that the statistical tests that were used should have been demonstrated to be generally accepted by the scientific community, but the court explained that because the opinion was based primarily on clinical experience and not solely on actuarial evidence, a *Frye* hearing was not required.[110] Under this approach, clinical opinion is admitted without a *Frye* threshold check, but research-based testimony that primarily relies on statistical techniques does encounter this evidentiary check.

These examples present a counterintuitive situation: research-based testimony, which often has more reliable foundations, is more likely to be excluded by the courts than less reliable clinical testimony. This situation creates considerable uncertainty regarding the standards for assessing the methodology of psychiatric and psychological testimony. On the one hand, one might argue that clinical mental health testimony should be subject to the same exacting scrutiny as framework evidence.[111] After all, as review of the psychiatric diagnostic process above shows, diagnosis has an empirical basis that can be conceptualized in methodological terms, not unlike research on eyewitness accuracy.

At the farthest extreme, some commentators have suggested that applying standards like *Daubert* literally would lead to the exclusion of most behavioral science opinion testimony.[112] On the other hand, mental health testimony is based on a reasonable analogue of what clinicians do

---

109. *See* People v. Miller, No. E035921, 2005 WL 768749, at *6 (Cal. Ct. App. Apr. 6, 2005); *see generally* David L. Faigman & John Monahan, *Standards of Legal Admissibility and Their Implications for Psychological Science*, *in* PSYCHOLOGICAL SCIENCE IN THE COURTROOM: CONSENSUS AND CONTROVERSY, *supra* note 51, at 3. Although California continues to be a "*Frye* state," a decision by the California Supreme Court employed the *Daubert* framework to buttress California's traditional approach, thus putting into doubt decisions such as *Miller*. *See* Sargon Enters., Inc. v. Univ. of S. Cal., 288 P.3d 1237, 1252 (Cal. 2012); *see generally* David L. Faigman & Edward J. Imwinkelried, *Wading into the* Daubert *Tide:* Sargon Enterprises, Inc. v. University of Southern California, 64 HASTINGS L.J. 1665, 1667 (2013) ("Although *Sargon* does not fully incorporate *Daubert* into California's Evidence Code, *Sargon* signals the court's interest in testing those waters.").

110. *Miller*, 2005 WL 768749, at *3, *5.

111. *See* FED. R. EVID. 702 advisory committee's note to 2000 amendments (citing Watkins v. Telsmith, Inc., 121 F.3d 984, 991 (5th Cir. 1997)) ("[I]t seems exactly backwards that experts who purport to rely on general engineering principles and practical experience might escape screening by the district court simply by stating that their conclusions were not reached by any particular method or technique.").

112. *See* Gottesman, *supra* note 102, at 875–76 ("[T]he Court's [*Daubert*] opinion read

every day—generally a probabilistic, provisional, and dynamic form of hypothesis testing—and in principle no different from other forms of medical diagnosis. It is, after all, on this basis that life-and-death decisions are made routinely in hospitals and other treatment settings.[113] Indeed, some courts have thought it appropriate to consider psychiatric diagnosis and functional assessment as "inductive" skills, based on experience and individualized to the case at hand, and thus not subject to the same scrutiny as "deductive" reasoning from group to individual.[114] This has been manifested in many jurisdictions as the so-called "opinion rule," which operates as an exception to the ordinary rules of admissibility.[115] As many commentators have found, however, this is an approach fraught with danger, and one that does not hold up under close scrutiny.[116] A closer look at the elements of G2i inference inherent in diagnosis and functional assessment may help to resolve this question.

B. *G2i Analysis of Admissibility*

We structure our analysis of admissibility of psychiatric testimony in light of G2i considerations according to the categorization developed by Faigman, Monahan, and Slobogin.[117] Their approach draws on elements similar to those that were considered by *Daubert* and its progeny,[118] and encourages judges to focus particularly on four considerations in making admissibility decisions regarding scientific

---

literally would dictate the end of the receipt of psychiatric and psychological testimony in federal courts.").

113. Psychiatrists routinely make decisions on admission, treatment, and discharge of patients who are suicidal or manifest homicidal ideation toward third parties.

114. *See* David L. Faigman, *Admissibility Regimes: The "Opinion Rule" and Other Oddities and Exceptions to Scientific Evidence, the Scientific Revolution, and Common Sense*, 36 SW. U. L. REV. 699, 709 (2008).

115. *See generally id.* at 707 ("[E]xpert opinion that is based on experience, and which claims no pretensions to scientific exactitude, does not receive this close evidentiary review. . . . This practice, one employed explicitly or implicitly in many jurisdictions, is often referred to as the 'opinion rule.'").

116. *See, e.g.*, D.H. Kaye, *Choice and Boundary Problems in* Logerquist*,* Hummert*, and* Kumho Tire, 33 ARIZ. ST. L.J. 41, 68 (2003) (concluding that the "broadly written exemption of 'personal knowledge' [i.e., knowledge based on the expert's own research and observations] from the demands ordinarily placed on scientific evidence is too generous"); David L. Faigman, *Embracing the Darkness:* Logerquist v. McVey *and the Doctrine of Ignorance of Science Is an Excuse*, 33 ARIZ. ST. L.J. 87, 90–91 (2003) ("If the expert is testifying to facts within his or her personal knowledge, then cross-examination is the great engine for discovering truth. . . . [But] if he or she is testifying on the basis of other people's experiences and opinions, then general acceptance must be demonstrated to the court. . . . [This] is not an approach well-tailored to obtaining reliable evidence.").

117. Faigman, Monahan & Slobogin, *supra* note 3, at 440 (listing "five criteria that are consistently identified as necessary considerations in determining the admissibility of scientific expert testimony").

118. *See supra* notes 46–49 and accompanying text.

evidence: (1) Relevance or "Fit"; (2) Helpfulness; (3) Reliability and Validity; and (4) Unfair Prejudice.[119] We next consider each element's application to both framework evidence and diagnostic evidence, especially in regards to testimony about diagnosis and functional assessment.

1. Relevance or "Fit"

As noted above, relevance refers to the match or fit between the evidence being proffered and the legal issue at hand.[120] The concept encompasses both legal and empirical concerns. Specifically, this inquiry considers whether the evidence that is being offered is relevant to the legal issue in contention and is based on research methods that relate to the factual question at hand. Although evidence concerning either legal or empirical relevance can be introduced in support of framework or *diagnostic* testimony, the issue of legal fit most commonly arises when framework evidence is considered, and the issue of empirical fit is most often raised when diagnostic evidence is at issue.[121]

i. Framework Evidence on "Fit"

The legal relevance of testimony regarding psychiatric diagnoses will typically be determined by the statute or case law governing the proceeding. Examples in which the law relies on diagnoses as a threshold consideration include insanity defense claims,[122] civil commitment proceedings,[123] and guardianship hearings.[124] However, in addition to

---

119. Faigman, Monahan & Slobogin, *supra* note 3, at 440. Their original treatment of this matter also considered a fifth factor, "qualifications," which we do not explore here.

120. *See supra* Part III.A.1.

121. But see Faigman, Monahan & Slobogin, *supra* note 3, at 443–47, for a discussion of the exceptions to this statement.

122. *See, e.g.*, N.Y. Penal Law § 40.15 (McKinney 2014).

> In any prosecution for an offense, it is an affirmative defense that when the defendant engaged in the proscribed conduct, he lacked criminal responsibility by reason of mental disease or defect. Such lack of criminal responsibility means that at the time of such conduct, as a result of mental disease or defect, he lacked substantial capacity to know or appreciate either:
>
> 1. The nature and consequences of such conduct; or
> 2. That such conduct was wrong.

Id.

123. *See, e.g.*, N.Y. Mental Hyg. Law § 9.37(a) (McKinney 2014) ("The director of a hospital, upon application by a director of community services or an examining physician duly designated by him or her, may receive and care for in such hospital as a patient any person who, in the opinion of the director of community services or the director's designee, has a mental illness for which immediate inpatient care and treatment in a hospital is appropriate and which is likely to result in serious harm to himself or herself or others.").

124. *See, e.g.*, Mental Hyg. § 81.02(b).

> The determination of incapacity shall be based on clear and convincing evidence and shall consist of a determination that a person is likely to suffer harm because:

cases in which the relevance of psychiatric diagnosis is embodied in the operative legal rule, courts may find such testimony material because it assists in the assessment of the parties' claims.[125] Thus, for example, even though emotional harms actions generally do not require testimony regarding the presence of a particular diagnosis resulting from the allegedly tortious act, such evidence may assist the factfinder in assessing the plausibility of a plaintiff's claims. Signs and symptoms that fit into an established syndrome may be more credible manifestations of harm than allegations of symptoms that match no known pattern, and hence be relevant to the proceeding.

A general challenge to the "fit" of evidence regarding the framework for psychiatric diagnoses was voiced by Stephen Morse, who argued that mental health experts should not testify about whether a person has a mental disorder or is "normal."[126] Among other objections, Morse claimed that mental disorder diagnostic categories are a poor match for the law's needs because they are overinclusive (or, in the terms discussed above, they have high sensitivity but low specificity).[127] In his view, the diagnostic categories encompass a set of symptoms much broader than the "crazy behaviors" that are legally relevant.[128] It is an open question (and one that relies in part on value judgments about how inclusive disorder criteria should be) as to whether these criteria are in fact "too" inclusive, but let us consider the extent to which an "overinclusiveness" objection undercuts the relevance of psychiatric diagnosis for legal purposes.

In the clinical practice of medicine, an overinclusive test may be appropriate, especially when the purpose of the test is screening. Indeed, inclusivity (i.e., sensitivity) is merely one component of accuracy, and it is always in tension with specificity. The question of where to define a diagnostic threshold is common to many areas of medicine. In the clinical realm, it is often the case that a first-pass screening test sets a

---

1. the person is unable to provide for personal needs and/or property management; and
2. the person cannot adequately understand and appreciate the nature and consequences of such inability.

Id.

125. Kennedy v. Municipality of Anchorage, 305 P.3d 1284, 1288 (Alaska 2013) (stating that, "for example, psychiatric records may point to sources of emotional distress other than the defendant's conduct").

126. Stephen J. Morse, *Crazy Behavior, Morals, and Science: An Analysis of Mental Health Law*, 51 S. CAL. L. REV. 527, 604 (1978) ("These decisions should not and cannot be totally dependent on scientific categories that may serve other purposes, and experts should not testify about whether an actor suffers from a mental disorder or even about whether the actor is normal.").

127. *See id.* at 605.

128. *Id.*

relatively low threshold for inclusion, erring on the side of high sensitivity, thus allowing positive results to be followed up with more stringent confirmatory tests that can rule-out false positives. For example, the screening test for tuberculosis is a skin test measuring a person's immune reactivity to tuberculin.[129] For most people, a "positive" test is defined as more than fifteen millimeters of redness and swelling at the injection site. However, this measurement does not reflect an absolute truth; it has merely been determined to be the optimal threshold for the first iteration of the test, after which a patient with a positive test must be further examined with targeted physical examinations or diagnostic testing (e.g., x-rays, blood tests).[130]

Similarly, mental disorder criteria that are somewhat overinclusive from a legal perspective (e.g., criteria for schizophrenia that identify a wider spectrum of cases than the law would choose to excuse from legal responsibility) may still have advantages in their use for legal purposes. Such criteria will be more sensitive in identifying persons who may qualify for the legal finding at issue (e.g., criminal non-responsibility), albeit less specific in their focus. Thus, they are less likely to exclude qualified persons. However, just as clinical screening tests serve merely as a trigger for further investigation, a mental disorder diagnosis does not settle the legal question on its own but acts as a threshold criterion for considering the subsequent functional arguments (e.g., whether a defendant lacked substantial ability to understand the wrongfulness of his behavior). Concerns about the overinclusiveness of diagnostic criteria for purposes of legal fit are thus mitigated by the recognition that actual impact on a person's capacity to function must also be demonstrated, which is a more specific determination.

With regard to framework evidence concerning functional abilities—the primary focus of legal inquiry—questions regarding the relevance of evidence that establishes a basis for determining functional ability will often be easily resolved by resorting to the legal standard at issue in the case. Functional determinations in insanity defense cases (e.g., did the defendant lack substantial ability to appreciate the wrongfulness of his behavior?),[131] commitment proceedings (e.g., is the respondent likely to represent a danger to himself or others?),[132] and guardianship hearings (e.g., does the alleged incompetent person lack the ability to manage his own affairs?)[133] are all examples on this point.

129. *See* Robin E. Huebner, Maybelle F. Schein & John B. Bass, Jr., *The Tuberculin Skin Test*, 17 CLINICAL INFECTIOUS DISEASES 968, 970 (1993).

130. *Id.*

131. *See, e.g.*, N.Y. PENAL LAW § 40.15 (McKinney 2014).

132. *See, e.g.*, N.Y. MENTAL HYG. LAW § 9.37(a) (McKinney 2014).

133. *See, e.g.*, MENTAL HYG. § 81.02(b).

Framework evidence in these cases would demonstrate the extent to which psychiatric evaluations can address these functional considerations.

Insofar as testimony about a party's functional abilities relates directly to the legal question at issue, any concerns expressed about the fit between psychiatric diagnosis and legal categories do not come into play. Under such circumstances, the psychiatrist's direct observations of the subject's behavior would have relevance under the applicable law. In effect, the expert would simply be providing descriptive evidence regarding the observed functional abilities in question, which is distinct from testimony that is prescriptive or inferential.

ii. *Diagnostic* Evidence on "Fit"

It is useful to approach the subject of relevance or fit of *diagnostic* evidence by taking into account the two kinds of fit described above: legal fit and empirical fit. As noted in the previous section, legal fit corresponds ordinarily with psychiatric framework evidence.[134] The substantive law establishes the relevance of particular empirical categories. Thus, for example, if "substantial lack of control" is relevant under applicable law, psychiatric diagnostic categories that indicate that certain groups of people (e.g., people with schizophrenia or intellectually disabled) typically have lowered "self-control," those psychiatric categories—or empirical frameworks—would legally fit. However, the question whether a particular defendant qualifies as an instance of a legally relevant empirical framework is a *diagnostic* question. At the diagnostic evidentiary level, empirical fit is the more salutary concern.

This issue turns out to be more complicated than it might first appear. Indeed, the question boils down to whether the psychiatric diagnosis is based on research that generalizes to the evidentiary diagnosis at issue in the case. Returning to the example of self-control, which is relevant in a variety of legal contexts, courts must ask whether the psychiatric diagnostic concept fits the legal diagnostic concept. This should require courts to examine the research basis for the psychiatric diagnostic category to determine whether it relates to the law's understanding or purposes. For instance, if all of the psychiatric research was done on non-incarcerated populations, the psychiatric diagnosis might not fit the evidentiary diagnostic issue in a case involving an incarcerated defendant. It is not that, for example, the diagnosis that the individual is "intellectually disabled" is diagnostically inaccurate as a psychiatric matter; rather, it is that the data indicating that intellectually disabled

134. *See supra* note 121 and accompanying text.

individuals who are not incarcerated have lowered self-control does not empirically fit the legally relevant *diagnostic* issue: will this intellectually disabled individual have lowered self-control when incarcerated?[135]

There might be an even more basic disconnect between the data on which psychiatric diagnoses are based and the legal issues in dispute. Continuing with the self-control example, suppose the psychiatric diagnosis is "intellectual disability."[136] At the framework level, this diagnosis is associated with a "lack of self-control." From the law's perspective, there are two diagnostic issues presented. First, whether the defendant is indeed "intellectually disabled." Psychiatrists using the *DSM* have sound criteria by which to make that assessment about individuals. Second, the court might also be interested to know whether an individual defendant who is intellectually disabled lacks self-control. It is not clear that psychiatric science has progressed far enough to support an evidentiary *diagnostic* statement to that effect. In this regard, there is a lack of empirical fit between the psychiatric diagnosis of intellectual disability and the legal *diagnostic* judgment that a particular person lacks self-control, because not all of those diagnosed as intellectually disabled will be *diagnosable* as lacking self-control.

2. Helpfulness (i.e., "Assist the Trier of Fact")

Even relevant expert evidence will not be admissible unless it is likely to help the finder of fact in understanding other elements of the evidence in the case or resolving a fact at issue.[137] But modern rules of evidence take a relatively permissive approach to the helpfulness criterion: The subject matter of an expert's testimony need not be beyond the ken of the average juror—the expert's testimony must merely provide sound assistance to the jury's factfinding role.[138]

---

135. For a discussion of empirical fit in the context of a case, see Faigman, Monahan & Slobogin, *supra* note 3, at 442 (discussing United States v. Carmel, 801 F.2d 997 (7th Cir. 1986)) ("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.").

136. Discussed *supra* notes 26–28 and accompanying text.

137. *See* Fed. R. Evid. 702(a); *see also* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 591 (1993) ("[Federal Rule of Evidence 702] further requires that the evidence or testimony 'assist the trier of fact to understand the evidence or to determine a fact in issue.'").

138. *See* Harvey Brown & Melissa Davis, *Eight Gates for Expert Witnesses: Fifteen Years Later*, 52 Hous. L. Rev. 1, 32 (2014). "[C]ourts generally resolve the issue in favor of admissibility, perhaps because if the testimony does not particularly aid the jury it may not particularly harm the opposing party either." *Id.* (citing *United States v. Lamarre*, 248 F.3d 642, 648 (7th Cir. 2001), which held that the trial judge erred in excluding an expert's testimony "merely because it overlap[ped] with matters within the jury's experience").

i. Framework Evidence on Helpfulness

Psychiatric framework evidence can be helpful when it allows laypersons to interpret what otherwise might appear to be random bits of data. For example, a report that two weeks prior to the commission of a crime the defendant had been talking rapidly and sometimes disjointedly, had been sleeping little, had engaged in promiscuous sexual activity, had spent a great deal of money, and had exhibited an elevated sense of his own abilities, might seem to many people as an odd but meaningless assortment of behaviors. However, a psychiatric expert's testimony that this cluster of behaviors is characteristic of a manic episode can be helpful to a judge and to jurors in putting that information into a meaningful context. Framework testimony of this sort can also be helpful in setting the stage for later diagnostic testimony. It might support the inference that the defendant himself suffers from bipolar disorder and was experiencing a manic episode at the time of the crime. Such testimony alternatively can lay the foundation for a demonstration that the defendant does *not* have a psychiatric disorder if, for instance, the pattern of signs and symptoms displayed by the defendant does not correspond to a known syndrome.

ii. *Diagnostic* Evidence on Helpfulness

When psychiatric evidence addresses the threshold question of the presence of a psychiatric disorder, it is likely to be helpful to the finder of fact because most laypersons lack the skills to make psychiatric diagnoses on their own. In principle, exceptions to that rule may occur when the factfinder can witness a party's impairment directly and independently draw the appropriate conclusions (e.g., when a respondent to a guardianship petition alleging the person is suffering from dementia takes the witness stand and appears disoriented, confused, and forgetful), or when framework testimony has clearly described the signs and symptoms associated with a given disorder (e.g., Alzheimer's disease). Even in these cases, however, courts are unlikely to exclude as unhelpful psychiatric testimony regarding diagnoses, since the stamp of authority conferred by an expert's testimony may reduce residual uncertainty about the conclusions to be drawn. Exclusion may be more likely for claims that are based on behavioral syndromes not generally recognized as mental disorders (e.g., battered woman's syndrome), if framework testimony has clearly described its components and the finder of fact is capable of ascertaining whether it is likely to be present in the case.[139]

139. *See, e.g.*, State v. Ritt, 599 N.W.2d 802, 811 (Minn. 1999) ("Testimony on battered woman syndrome is limited to a description of the syndrome's general nature and the expert is not allowed to testify whether a particular defendant or witness suffers from the syndrome because the

Psychiatric evidence that speaks specifically to the functional impairment at issue in the case (e.g., lack of ability to form a specific intent or inability to manage one's affairs) is likely to be deemed as helpful, assuming legal rules do not bar its introduction for other reasons.[140] However, when testimony about functional abilities is framed in terms of the ultimate legal standard, it may be excluded as trenching on the prerogative of the finder of fact. The most notable example is embodied in the federal insanity defense statute, which bars the introduction of expert testimony about a defendant's ability to "appreciate the nature and quality or the wrongfulness of his acts," which is the standard for criminal responsibility.[141] Even in jurisdictions without a statutory bar to the admission of testimony that addresses the ultimate legal issue of a specific functional capacity of the defendant or plaintiff, judges retain the discretion to block the introduction of such testimony as going beyond the expertise of the witness or invading the province of the jury.[142]

3. Reliability and Validity of Methods and Principles

Since psychiatric evidence will be relevant to a large number of cases, and in principle it is likely to be helpful to the finder of fact when mental state and behavioral capacities are at issue, questions regarding admissibility will often come down to issues related to reliability and validity. Are the psychiatric categories that are being described valid (i.e., "real") entities that can be distinguished reliably from other mental disorders or the absence of a mental disorder, and do they impact the functional abilities in question? The *Daubert* Court identified four factors related to the reliability and validity of expert scientific testimony that judges should consider when determining the admissibility of such evidence: (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

---

expert testimony may be perceived as evidence on the ultimate issue of guilt or innocence . . . or as an 'unwarranted 'stamp of scientific legitimacy' to the testimony.' ") (internal citation omitted); *see also* State v. Laprade, 958 A.2d 1179, 1186–87 (Vt. 2008) (upholding trial court's allowance of prosecutorial use of the syndrome to explain the woman's recantation when the expert "confine[d] her testimony to her general knowledge, and [did] not comment on the parties or the specific facts of the case.").

140. *See, e.g.*, Clark v. Arizona, 548 U.S. 735, 774 (2006) (refusing to invalidate an Arizona statute that prohibited expert testimony about the defendant's diagnosis, or what the Court called "mental-disease" evidence).

141. 18 U.S.C. § 17 (2012); *see also* Fed. R. Evid. 704(b).

142. As discussed above, *Daubert* declares that judges have the responsibility to examine the methods and principles that underlie a proffered expert's testimony. *See supra* note 37 and accompanying text.

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.[143] *Kumho Tire Co.* 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."[144] These factors were not meant as a simple checklist, but instead describe the sorts of criteria that scientists themselves might consider in assessing scientific claims. In that way, they ought to be evaluated separately and together, as part of the overall evaluative process of assessing reliability and validity.

i. Testing

Testability of evidence goes to the heart of the scientific basis for an expert's conclusions. Only propositions that are testable can be considered subject to the scientific method or capable of being empirically verified or disproved. Expert testimony that is based on untestable assertions or generalized "experience" may be admissible in some cases (e.g., the conclusion of an art historian regarding the painter who produced a particular work of art—"In my opinion, this was painted by Renoir"—and its value—"Based on my experience, I would estimate this painting to sell at auction for between $4 million and $6 million").[145] However, this type of testimony is not scientific testimony and, hence, needs to be judged by standards other than those laid out in *Daubert* and related cases at the state level.[146] On the other hand, psychiatric testimony—as a type of medical testimony—is generally presumed to have a scientific basis and, insofar as possible, must be based on testable propositions.

a. Framework Evidence on Testing

As applied to framework testimony about psychiatric diagnosis, testability requires clear-cut criteria for defining disorders and determin-

143. *Daubert*, 509 U.S. at 592–94.

144. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 152 (1999).

145. It is important to note, however, that even when experts rely on their own experience, they still must explain how their experience is related to or truly supports their proposed testimony. *See* FED. R. EVID. 702 advisory committee's note to 2000 amendments ("If the witness is relying solely or primarily on experience, then the witness must explain how that experience leads to the conclusion reached, why that experience is a sufficient basis for the opinion, and how that experience is reliably applied to the facts. The trial court's gatekeeping function requires more than simply 'taking the expert's word for it.'") (citing Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1319 (9th Cir. 1995)).

146. *See* Kumho Tire Co., 526 U.S. at 150 ("[T]here are many different kinds of experts, and many different kinds of expertise. . . . [W]e can neither rule out, nor rule in, for all cases and for all time the applicability of the factors mentioned in *Daubert*, nor can we now do so for subsets of cases categorized by category of expert or by kind of evidence.").

ing when they are present. A second component of Morse's challenge to the introduction of psychiatric diagnosis in court is the allegation that one cannot "conclusively verify" whether someone has a mental disorder—i.e., the constructs of mental disorders are not testable.[147] Therefore, Morse argues, because "there is no such thing as an independently 'correct' or 'incorrect' mental health diagnosis," the role of experts should be limited to describing their observations of a person's behavior.[148] He further maintains that they should not be permitted to offer opinions about the normality or pathological nature of that behavior, or to place a diagnostic label on it.[149] Because verification is rare in psychiatric diagnosis, and given the absence of biological measures, we suggest that conclusive verification of a litigant's psychiatric diagnosis is neither a useful nor a legally relevant standard.

Psychiatric diagnoses are customarily used in clinical practice for highly consequential decisions, often as consequential as the determinations made in court. Judgments about whether the data supporting the validity of a given diagnosis are sufficient for the diagnosis to be admissible for legal purposes must be made in light of the specific data on which the construct is based. In that regard, the task force working on the most recent revision of the *DSM* identified a set of validators to which they would look to determine the validity of a proposed diagnosis.[150] Under the task force's proposed evaluative method for validity of a diagnosis, the presumed validity of a proposed diagnosis would be bolstered by showing that persons diagnosed with a given criteria set shared groups of common traits.[151] These include: (1) antecedent validators (familial aggregation; socio-demographic and cultural factors; environmental risk factors; prior psychiatric history); (2) concurrent validators (cognitive, emotional, temperament, and personality correlates; biological markers, e.g., molecular genetic markers, neural substrates) and patterns of comorbidity; and (3) predictive validators (diagnostic stability, course of illness, response to treatment).[152] Similar standards are generally applied to other medical diagnoses. In the end, then, the question should not be whether diagnostic categories meet an impossibly high standard of absolute validity (e.g., "conclusive verification"); the evidence supporting a given psychiatric diagnosis will admit

147. *See* Morse, *supra* note 126, at 607.

148. *Id.* at 611–12 (positing than an expert witness "need not and should not report conclusions about mental disorder, abnormality, or even craziness; these are legal determinations for the judge or jury" and that "[i]t is far more precise and useful to the judge or jury if the expert simply describes his observations of behavior").

149. *See id.* at 607–11.

150. *See* Kendler et al., *supra* note 77.

151. *See id.*

152. *See id.*

of degrees. The question ought to be whether evidence from the available validators is sufficiently strong to be of use for the particular context in which the diagnosis is being applied. Thus, when the admissibility of a psychiatric diagnosis is called into question on the basis of its testability, it is imperative to consider the bases on which the diagnostic construct rests. There can be no one-size-fits-all approach to the testability of mental health evidence.

Framework evidence about functional abilities is less theoretically fraught than framework evidence regarding diagnoses but nonetheless potentially challenging to evaluate. Complicating the analysis in many cases is the paucity of data on the relationship between psychiatric disorders and many legally relevant functional impairments. Unfortunately, systematic examination of the functional abilities that are likely to be at issue in court began only within the last several decades, and the aggregate data from which conclusions can be drawn are often extremely limited.[153] The extent of available framework evidence also varies by the temporal focus of the evaluation. Retrospective assessments (e.g., what was defendant's state of mind at the time of a crime?) are the least well-validated, in part because of the difficulty of gathering confirmatory data;[154] predictive assessments (e.g., how well will this parent in a custody dispute raise his children if granted custody?) have variable amounts of framework evidence, with the best-developed body of data applying to violence risk assessments.[155] In contrast, current functioning assessments (e.g., what is this alleged incompetent person's capacity to manage her financial affairs?) likely have the largest amount of framework evidence, although even here there is considerable variation.[156] Thus, experts' conclusions are likely to be based on extrapolations from limited data sets, a body of documented clinical observations (e.g., case reports of delusionally motivated behaviors in schizophrenia), and the expert's own clinical experiences.

Although most psychiatric assessments of functional impairment today derive from clinical examination supplemented by collateral information, the last several decades have seen a proliferation of structured

153. *See* THOMAS GRISSO, EVALUATING COMPETENCIES: FORENSIC ASSESSMENTS AND INSTRUMENTS 80 (2d ed. 2003).

154. For a discussion of retroactive assessments, see generally AM. PSYCHIATRIC PUBL'G, RETROSPECTIVE ASSESSMENTS OF MENTAL STATES IN LITIGATION (Robert I. Simon & Daniel W. Shuman eds., 2002).

155. *See, e.g.*, Stefanía Ægisdóttir et al., *The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction*, 34 COUNSELING PSYCHOLOGIST 341 (May 2006).

156. *See* GRISSO, *supra* note 153, at 61–62 ("In summary, the real world may not allow one to test the predictive or postdictive validity of FAIs [forensic assessment instruments].").

instruments that are designed to aid in these assessments.[157] These tools range from instruments for assessment of competence to stand trial or competence to consent to medical treatment, to parenting capacity, violence risk (including special instruments to assess likelihood of sexual offenses), criminal responsibility, and ability to manage tasks of daily living. Some tools merely structure the evaluation and leave it to the assessor to draw conclusions about the nature and degree of impairment. Other instruments provide a quantitative measure of impairment, sometimes with explicit cut-offs to indicate when the impairment has reached a legally relevant point. Use of such instruments in court is still relatively uncommon, although there is reason to believe that some classes of instruments are now used more often (e.g., violence risk assessment tools),[158] and some are even statutorily required in certain states (e.g., sexual offense recidivism risk assessment tools).[159] Psychologists, with their extensive training in the use of assessment instruments,[160] are more likely to employ these approaches than other types of mental health professionals.

When assessment instruments are used, among the questions that should be asked are whether there are data indicating the reliability and validity of the instrument and how strongly the data support the instruments' use. Relevant areas to explore here will include interrater reli-

157. *See generally id.*

158. VA. CODE ANN. § 37.2-903(C) (West 2014) ("If the Director and the Commissioner agree that no specific scientifically validated instrument exists to measure the risk assessment of a prisoner, the prisoner may instead be screened by a licensed psychiatrist, licensed clinical psychologist, or a licensed mental health professional . . . for an initial determination of whether or not the prisoner may meet the definition of a sexually violent predator."); *see* M. Doyle & M. Dolan, *Violence Risk Assessment: Combining Actuarial and Clinical Information to Structure Clinical Judgements for the Formulation and Management of Risk*, 9 J. PSYCHIATRIC & MENTAL HEALTH NURSING 649, 652–53 (2002).

159. § 37.2-903(B).

> Each month, the Director shall review the database and identify all such prisoners who are scheduled for release from prison within 10 months from the date of such review or have been referred to the Director by the Virginia Parole Board under rules adopted by the Board (i) who receive a score of five or more on the Static-99 or a similar score on a comparable, scientifically validated instrument designated by the Commissioner, (ii) who receive a score of four on the Static-99 or a similar score on a comparable, scientifically validated instrument if the sexually violent offense mandating the prisoner's evaluation under this section was a violation of § 18.2-61, 18.2-67.1, 18.2-67.2, or 18.2-67.3 where the victim was under the age of 13, or (iii) whose records reflect such aggravating circumstances that the Director determines the offender appears to meet the definition of a sexually violent predator. The Director may exclude from referral prisoners who are so incapacitated by a permanent and debilitating medical condition or a terminal illness so as to represent no threat to public safety.

160. This is partly due to the fact that psychologists have "cornered the market on clinical assessment." Amy Novotney, *Postgrad Growth Area: Assessment Psychology*, GRADPSYCH MAG., Mar. 2010, at 26, *available at* http://www.apa.org/gradpsych/2010/03/postgrad.aspx.

ability (do different evaluators using the instrument come up with similar scores or conclusions?), test-retest reliability (are the findings from the instrument likely to be stable over time?), content validity (does the instrument reflect the legally relevant constructs for the assessment being performed?), convergent validity (does the instrument yield similar results to other means of assessing this functional capacity?), and discriminant validity (how well does the instrument perform in distinguishing impaired from unimpaired groups?). In addition, decisions about admissibility should take into account whether the instrument has been shown to be valid for the class of persons of which the defendant or plaintiff is a member and for the setting in which the instrument is being used. Taken together, these two considerations encompass the notion of generalizability, that is, whether we can extrapolate from the people studied in the development and validation of the instrument to the current subject(s) and from the experimental situation to a real-world setting. For example, a violence risk prediction instrument validated exclusively on a sample of men may not be equally applicable to women.[161] In addition, even instruments with good psychometric characteristics (i.e., reliability and validity) will yield inaccurate results if not administered in a standardized fashion, and some instruments require formal training for reliable application.[162] Finally, even when assessment instruments yield a quantitative result, they may call for clinician judgment in drawing the ultimate, legally relevant conclusion (i.e., the defendant is not capable of cooperating with her attorney in her own defense), a step that may introduce considerations of reliability and validity of its own.

b. *Diagnostic* Evidence on Testing

In regard to the diagnosis and functional ability of a particular party in a case, the *Daubert* concept of testability is inherently problematic. By definition, validity and reliability require analysis of aggregate data. One cannot determine in a rigorous way the validity or reliability of a single judgment. Even inquiry regarding the accumulated experience of a given psychiatric expert is usually impossible, since data on the accu-

161. *See* Doe v. Sex Offender Registry Bd., 999 N.E.2d 478, 488 & n.10 (Mass. 2013) (vacating classification of female appellant as a Level 1 sex offender on the grounds that the criteria applied were derived entirely from studies of recidivism in male sex offenders).

162. *See, e.g.*, *The Hare PCL-R Training Program*, WITHOUT CONSCIENCE, http://www.hare.org/training/ (last visited Apr. 3, 2015) (advertising PCL-R training workshops for clinicians and professionals who conduct psychological assessments); *Evaluation of Risk for Violence Using the HCR-20 Version 3*, CONCEPT, https://ao208.infusionsoft.com/app/page/default-campaign-landing-page5 (last visited Mar. 30, 2015) (advertising training programs for the HCR-20, the "most commonly used violence risk assessment measure across 44 different countries").

racy of their diagnostic and functional determinations are almost never available. However, in place of outcome measures, courts can resort to process analysis. That is, to the extent that valid and reliable processes have been defined for the assessment in question (e.g., making a psychiatric diagnosis of depression by applying *DSM-5* criteria, or evaluating competence to stand trial of a criminal defendant with a standardized instrument developed and validated for the purpose), it is possible to ask about the extent to which the expert followed the appropriate process.

When it comes to establishing a standard psychiatric diagnosis, professional guidelines call for application of the relevant *DSM-5* criteria.[163] However, accurate determination of the presence of relevant signs and symptoms requires a careful, and often sensitive, clinical examination. The likely reliability and validity of diagnostic conclusions can be estimated, in part, by determining whether a thorough clinical evaluation was performed, with inquiry into the usual areas of investigation, including history of the presenting illness, past psychiatric history, social history (including educational and occupational histories), family history, medical history (including substance use and prescribed medications), and contemporaneous mental status examination.[164] Depending on the circumstances, data may also need to be gathered from records (e.g., past medical records) and collateral sources (e.g., family members). In addition, although currently used only in a minority of clinical evaluations, there are a growing number of structured diagnostic instruments that can be utilized.[165] In general, structured approaches achieve higher levels of reliability, but they are time-consuming, may be difficult to apply in forensic settings, and often raise questions of validity (i.e., whether the questions in the assessment schedule effectively capture the constructs embodied in the diagnostic criteria). If a testifying expert uses a structured diagnostic instrument, the court must consider the instrument's validity and reliability in general, as well as whether there may have been limitations on its likely accuracy in a particular case (e.g., an uncooperative evaluee or an untrained examiner).

Many commentators believe that the use of structured assessments will improve the quality of psychiatric testimony regarding functional impairments.[166] However, the promise of greater reliability and validity in psychiatric assessments depends on the quality of the instruments,

163. *See supra* Part III.B.1.

164. MELTON ET AL., *supra* note 85, at 43–67.

165. GRISSO, *supra* note 153, at 41–67.

166. *See* Charles L. Scott, *Believing Doesn't Make It So: Forensic Education and the Search for Truth*, 41 J. AM. ACAD. PSYCHIATRY & L. 18, 23–25 (2013) (noting that the "scientific literature . . . indicates significant support for the use of more structured methods" in forensic psychiatry).

their appropriate use, and correct interpretation of their results. Poorly designed measures or instruments that are used improperly will only confuse the factfinding process, as would conclusions reaching beyond the actual explanatory power of the instruments being used. Hence, whether these instruments have been correctly applied to individual cases depends on the data used in the respective case.

One final caveat regards an additional variable—the possibility of bias. When a litigant challenges a diagnosis based on reliability grounds, it may seem that diagnoses that rely on standard criteria (i.e., those based on the *DSM*) and are arrived at through usual diagnostic practices should generally be considered admissible, as the clinical scientific literature generally supports the reliability of those practices. However, the adversarial nature of the legal context necessitates further cautions about generalizing from data that would suggest good diagnostic reliability in clinical settings. Of course, there may be deliberate bias when an expert is consciously motivated to produce a result advantageous for his or her side. But some research supports the possibility that mental health experts may be unconsciously biased as well. In a study of experts in sexually violent predator cases who used a standardized rating instrument to measure psychopathy, prosecution experts produced psychopathy scores consistently higher than those of defense experts—i.e., the usual reliability of the instrument used was markedly reduced in the context of a legal contest.[167] The same research group has also found preliminary evidence that other forensic assessment tools may be subject to expert bias in adversarial contexts. For example, one study of the use of an actuarial risk assessment tool in sexually violent predator assessments found evidence for adversarial allegiance—i.e., risk scores from opposing evaluators differed in a direction that supported the party who retained their services.[168] Thus, reasons to suspect bias, conscious or unconscious, should be carefully examined.

ii. Error Rate

The concept of an error rate is intrinsically related to the testability of the underlying process on which expert testimony is based. No scientific process is perfectly accurate when repeated many times. Some degree of error is present in any analytic approach, which means that even if most outputs are accurate (i.e., the error rate is low), some out-

167. *See* Daniel C. Murrie et al., *Are Forensic Experts Biased by the Side that Retained Them?*, PSYCHOL. SCI., Aug. 2013, at 1, 7–8.

168. *See* Daniel C. Murrie et al., *Rater (Dis)agreement on Risk Assessment Measures in Sexually Violent Predator Proceedings: Evidence of Adversarial Allegiance in Forensic Evaluation?*, 15 PSYCHOL. PUB. POL'Y & L. 19, 48 (2009).

puts will be wrong. Error can be systematic (e.g., a set of diagnostic criteria omits a key symptom that is highly associated with the condition) or random (e.g., a psychiatrist occasionally forgets to ask about one or another of the diagnostic criteria and is left with an incomplete set of data from which to make a diagnostic judgment), and can affect both the reliability and the validity of a determination.

a. Framework Evidence on Error Rate

Not all mental disorder constructs show the same degree of reliability and validity. They can be quite heterogeneous with regard to rates of error. For instance, the classic disorder of schizophrenia has been well characterized after decades of study, and accordingly, its diagnostic reliability over time has been good, as demonstrated in studies beginning in the 1970s and continuing through the recent *DSM-5* field trials.[169] In contrast, "attenuated psychotic syndrome" ("APS") was proposed as a new disorder for the *DSM-5*, as a type of prodrome (i.e., an early set of symptoms preceding the full onset of disorder) for psychotic disorders such as schizophrenia.[170] Although some research supports its existence, a lack of broader sufficient empirical support for its status as a disorder led to its being included only in a section of *DSM-5* identifying conditions for future study.[171]

Reliability and validity must also be distinguished from one another, and keeping these concepts straight can help to parse the data on error rates. For example, the relatively new concept of APS was thought to have questionable validity, despite some preliminary data suggesting that its interrater reliability might be acceptable.[172] Some of those validity concerns had to do with real-world implementation issues, such as the potential for misdiagnoses (especially the likelihood of a substantial number of false positive diagnoses) or incorrect ideas about what the diagnosis indicated (such as that persons so diagnosed will inevitably develop a full-fledged psychotic disorder).[173] In the end, the conclusion was that more research was needed to support its status as a

169. *See* Regier et al., *supra* note 74, at 62 (reporting schizophrenia to be within the "good" kappa range of between 0.40 and 0.59).

170. *See id.* at 66 ("[A]ttenuated psychotic symptom syndrome disorder . . . ha[s] a substantial body of research supporting [its] inclusion in DSM-5, but [its] reliability could not be assessed with the available samples at large academic settings.").

171. *See* Ming T. Tsuang et al., *Attenuated Psychosis Syndrome in DSM-5*, 150 SCHIZOPHRENIA RES. 31, 31–35 (2013) ("The Workgroup decided to recommend the inclusion of Attenuated Psychosis Syndrome as a category in the appendix (Section 3) of DSM-5 as a condition for further study.").

172. *See* Regier et al., *supra* note 74, at 64, 66.

173. *See generally* Tsuang et al., *supra* note 171.

disorder, but not because of reliability problems.[174] In contrast, some diagnoses studied in *DSM-5* field trials did not achieve strong reliability, despite considerable agreement that they qualify as true mental disorders. For example, "alcohol use disorder" was found to have a middling interrater kappa value of 0.40, on the border of the "questionable" range.[175] However, few clinicians would dispute the idea that there exists a class of substance-related and addictive disorders. Despite uncertain diagnostic reliability measures for this class, then, its validity as a mental disorder is supported by the clinical community, and it continues to be recognized by the *DSM-5*.[176] Thus, while quantitative measures of reliability can be helpful in determining admissibility, they will need to be considered in light of other evidence regarding a particular disorder.

With regard to functional assessments, data typically are sparse to non-existent regarding rates of error for particular determinations (e.g., judgments of competence to waive *Miranda* rights). The situation is somewhat better when structured assessment instruments are used in these evaluations, since authors of instruments usually will have published data characterizing their validity and reliability.[177] However, there are often questions about the comparability of the population on which an instrument has been developed and tested to the person being evaluated in a given case that complicate the interpretation of such data. As noted above, the paradox in most courts today is that, because data on error rates are available for structured assessment instruments but not for clinical evaluations, the former, which may have higher levels of reliability and validity, tend to be scrutinized more closely than the latter, with which courts have grown familiar over the years.[178]

In making admissibility determinations, courts will need to assess whether rates of error in a diagnostic process or functional assessment—even if acceptable to the clinical community—are tolerable in the court-

174. *See id.* (stating that while the "reliability of [an APS] diagnosis is well established in academic and research settings, it was found to be less [reliable] in community and other clinical settings," and also that the "relationship between APS and other psychiatric conditions was unclear").

175. Regier et al., *supra* note 74, at 63. A kappa coefficient of 0.2 to 0.39 falls within the "questionable" range, whereas a kappa coefficient of 0.4 to 0.59 falls within the "good" range. *Id.* at 62.

176. *See* DSM-5, *supra* note 19, at 490–92.

177. *See, e.g.*, Marson et al., *supra* note 91, at 880–82.

178. *See* Daniel A. Krauss & Nicholas Scurich, *Risk Assessment in the Law: Legal Admissibility, Scientific Validity, and Some Disparities Between Research and Practice*, 31 BEHAV. SCI. & L. 215, 216, 219 nn.11–12 (2013) (stating that "courts simply have a long history of accepting clinical judgment expert testimony in a variety of contexts, including risk," and that "structured risk assessment approaches were more clearly . . . scientific, and thus amenable to evaluation under expert testimony standards").

room. Whereas clinicians faced with a patient in distress may choose to treat on the basis of a more-likely-than-not diagnostic determination (i.e., where the error rate could approach 49%), recognizing both the patient's need for assistance and the clinician's ability to modify the diagnosis over time as additional information becomes available (e.g., response to treatment), considerations in the courtroom may be substantially different. Finders of fact often are making one-time judgments that are not easily susceptible to alteration as new information appears. If psychiatric evidence is to be relied on for that purpose, courts might reasonably want to know that the likely rate of error is lower than might be tolerated in the clinic.

b. *Diagnostic* Evidence on Error Rate

Once a sufficiently valid and reliable framework for assessing diagnostic and functional characteristics has been established, the next step in the assessment of admissibility will turn on the extent to which general rates of error are likely to apply to the current evaluation. Even diagnostic criteria with strong indicators of validity may be applied in a haphazard fashion, yielding conclusions of questionable reliability. Put differently, there may be no question that schizophrenia is a valid diagnosis, but considerable doubt may exist as to whether *this* defendant suffers from it. Data regarding rates of error in the assessment of particular disorders may be helpful here, since some criteria are easier to apply reliably than others. For example, auditory hallucinations are a common symptom of schizophrenia that is capable of reliable determination.[179] On the other hand, the most recent revision of the *DSM* no longer gives special status in the diagnostic process for schizophrenia to the presence of "bizarre delusions."[180] Previously, bizarre delusions were considered particularly strong evidence of schizophrenia, and the presence of this symptom alone was—in combination with duration and impairment criteria—sufficient to make the diagnosis.[181] The *DSM-5*, however, eliminated the special status of bizarre delusions because clinicians in practice had difficulty reliably distinguishing bizarre from nonbizarre delusions.[182] However, even if a set of diagnostic or functional criteria can be shown to be susceptible to reliable application, it may still be the case that it was employed in a given case in a fashion that was

179. *See* A.Y. Tien, *Distributions of Hallucinations in the Population*, 26 SOC. PSYCHIATRY & PSYCHIATRIC EPIDEMIOLOGY 287, 287 (1991).

180. *See* AM. PSYCHIATRIC ASS'N, HIGHLIGHTS OF CHANGES FROM DSM-IV-TR TO DSM-5 2–3 (2013), *available at* http://www.dsm5.org/Documents/changes%20from%20dsm-iv-tr%20to%20dsm-5.pdf.

181. *See id.*

182. *See id.*

likely to impair its reliability. Overly brief clinical examinations, uncooperative evaluees, failure to examine past treatment records, and unavailability of collateral informants can all limit the reliability with which a diagnosis can be made. As is evident, these are fact-specific determinations regarding the degree of reliance on a diagnostic judgment or functional assessment in a given case.

In principle, determinations of admissibility might be facilitated by knowing the error rates of individual experts. Some psychiatrists, for example, may be better trained or simply more meticulous about the conduct of their evaluations. Experts may differ in the degree to which they remain current with the research and clinical literature, and hence the extent to which they are able to apply contemporary understandings to the diagnostic or functional assessment process. The result may be consistent differences in the accuracy (i.e., false positive and false negative rates) with which determinations are made. However, it is almost never the case today that such data are available. For judgments regarding past events (e.g., mental state at the time of a crime) and contemporary status (e.g., competence to stand trial), there is no "gold standard" determination against which an expert's judgment can be compared, and thus error rates cannot be calculated. For predictive judgments (e.g., likelihood that a sexual offender will recidivate), systematic follow-up is usually lacking and interventions undertaken in response to the prediction often preclude assessment of its accuracy (e.g., a sexual offender who is identified as likely to commit subsequent offenses may be involuntarily committed to a treatment facility, leaving the accuracy of the judgment incapable of being assessed because commitment prevents the predicted offenses from occurring).

Although consideration of error rates associated with particular experts is unlikely to be possible in most cases, *Daubert* provided some additional guidance as to how courts might nonetheless assess the likelihood of error. By indicating that "in the case of a particular scientific technique, the court ordinarily should consider . . . the existence and maintenance of standards controlling the technique's operation," the Court pointed to the importance of standardized approaches to assessment based on empirical data, where available, and on professional standards, where data are lacking.[183] One of the advantages of *DSM* for the courts, with its structured lists of diagnostic criteria, is that it provides a clear standard by which an expert's diagnosis can be judged. Assuming framework evidence has established to the court's satisfaction the validity and reliability of a particular set of diagnostic criteria, the expert's

183. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 594 (1993).

reliance on those criteria provides some assurance of the accuracy of the diagnosis.

An additional consideration is the reliability with which the expert gathered the information on which a determination is made of whether a criterion has been met. As an example, one of the criteria for a diagnosis of major depression is "markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day."[184] The reliability of this determination can be reduced by an evaluator neglecting to ask the question, failing to ascertain whether a reported lack of pleasure in a particular activity (e.g., watching football games) extends to all or almost all other activities, or not inquiring whether the person's diminution in interest lasts most of day and is present nearly every day. Of course, the failing may not be the expert's. The party being evaluated may avoid responding to a question or respond in an ambiguous manner that cannot be clarified with follow-up questions. Or neither party may be at fault, but the setting in which the evaluation takes place is sufficiently noisy or otherwise distracting that the evaluee cannot focus on the question and provides inaccurate information, or the examiner cannot hear the response. When considering the quality of the evaluation, a court may need to consider whether professional standards were adhered to in its conduct. For example, did it take place in a quiet, private space, with ample time for questions to be asked and probed, and was the evaluator systematic about the process of gathering the relevant information?[185]

Many of the same considerations apply to assessments of functional impairment. Assuming the clinical assessment process or the structured assessment instrument that was used is sufficiently valid and reliable to pass the framework analysis, courts will then need to assess whether the expert conformed to accepted standards in the conduct of the evaluation, including the use of any instruments involved. In some respects, this is a more challenging determination than the comparable assessment of testimony about diagnosis because there is no *DSM-5* equivalent for functional capacities.[186] Beyond the operative legal standards, which can be stated at a high level of abstraction (e.g., "the defendant, as a result of a severe mental disease or defect, was unable to appreciate the nature and quality or the wrongfulness of his acts"[187]), there is typically no defini-

184. DSM-5, *supra* note 19, at 125.

185. For summary of the "nature and method of forensic assessment, see MELTON ET AL., *supra* note 85, at 43–67.

186. The field of functional capacity assessment is at a sufficiently early stage that development of a comprehensive consensus document detailing the basis for judgments of functional capacities would lack sufficient empirical grounding and hence would be premature.

187. 18 U.S.C. § 17 (2012).

tive source of criteria to be applied. Hence, in addition to assessing whether the examination process comports with accepted professional standards (organizations of forensic professionals have produced ethics guidelines for many—though not all—forensic evaluations),[188] the court must assess the reasonableness of the criteria applied to the determination of the presence and degree of functional impairment. Some guidance with regard to the latter will come from professional guidelines and textbook authors, both attempting to distill best practices of the field, but there will usually be less guidance available for the court than in the case of standard psychiatric diagnoses.

iii. Peer Review and Publication

Publication of research findings in peer-reviewed journals is a basic expectation in much of academia, including psychiatry (and the other health and mental health professions). This includes studies of the validity and reliability of diagnostic and functional assessments. Journals conduct peer review by sending submitted manuscripts to reviewers who the journals consider knowledgeable about the topic being addressed. Reviewers' comments are taken into consideration by the editors in determining whether to accept or reject the manuscript, or to request revisions prior to re-review.[189] The quality of a peer-review process depends on the degree of knowledge of the pool of reviewers on which a journal can draw and how carefully the reviewers who are chosen consider the manuscript. Since journal editors can elect to ignore some or all of a reviewer's objections, their commitment to peer review and their personal knowledge of the topic under consideration play an important role as well. Although imperfect, publication in peer-reviewed journals is relied upon for decisions about promotion, grant awards, and allocation of other resources.

Not every type of document that may be useful to the court in assessing the admissibility of a psychiatric evaluation, though, will appear in peer-reviewed literature. For instance, guidelines produced by professional organizations, having gone through a consensus-development process, will often not be subject to subsequent peer review. Moreover, the quality of peer-reviewed journals is not equal. Some journals—usually those with higher "impact factors"[190]—are generally recognized as more rigorous than others. The proliferation of professional journals

188. *See generally Ethics Guidelines for the Practice of Forensic Psychiatry*, AM. ACAD. PSYCHIATRY & L., http://www.aapl.org/ethics.htm (last visited Aug. 1, 2014).

189. *See* Ray Spier, *The History of the Peer-Review Process*, 20 TRENDS IN BIOTECHNOLOGY 357 (2002).

190. *See* Eugene Garfield, *The History and Meaning of the Journal Impact Factor*, 295 J. AM. MED. ASS'N 90, 90–92 (2006).

in recent years, driven in part by commercial models of online publication that rely on payments from authors to cover the cost of production, have diluted the overall quality of the published literature and heightened the disparities across journals.[191] Moreover, even some research published in the best peer-reviewed journals is ultimately shown to be wrong.[192] Hence, while peer review for publication purposes can be a helpful indicator of the quality of the database that underlies an expert's reasoning, it is not a *sine qua non* of admissibility.[193]

Finally, it is worth noting briefly that peer review occurs in scientific fields in many ways that extend beyond the referee process endemic in academic journals. For instance, research that is supported by public or private grants has often received extensive formal peer review. Books published by university presses similarly are peer reviewed, often both before a contract is offered to the author and again when a final manuscript is submitted. And, of course, scientists continue to review and comment on research long after its publication date through analyses that might be found in blogs, websites dedicated to such commentary, letters to the editor, or subsequent research papers.

a. Framework Evidence on Peer Review and Publication

Criteria for standard psychiatric diagnoses appear in the *DSM-5*. Although not a peer-reviewed publication in the usual sense, the *DSM* development process integrated reviews by many experts and clinicians. Work groups were appointed for each substantive area of psychiatric diagnosis (e.g., schizophrenia and other psychotic disorders), comprising experts in that area, to develop proposals for changes in, additions to, or deletions of diagnostic categories.[194] Their proposals were posted online, with public comments solicited.[195] In some cases, the effect of the proposed changes was tested in field trials, where the reliability of resulting diagnostic determinations was assessed. Data from the trials were subse-

191. *See* Gina Kolata, *Scientific Articles Accepted (Personal Checks, Too)*, N.Y. TIMES, Apr. 8, 2013, at A1, *available at* http://www.nytimes.com/2013/04/08/health/for-scientists-an-exploding-world-of-pseudo-academia.html?hp ("[S]ome researchers are now raising the alarm about what they see as the proliferation of online journals that will print seemingly anything for a fee. They warn that nonexperts doing online research will have trouble distinguishing credible research from junk.").

192. *See, e.g.*, Haruko Obokata et al., *Retraction: Bidirectional Developmental Potential in Reprogrammed Cells with Acquired Pluripotency*, 511 NATURE 112, 112 (2014) (retracting an article in its entirety because multiple errors—some of which even rose to the level of misconduct—were found to "impair the credibility of the study as a whole").

193. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593 (1993).

194. DSM-5, *supra* note 19, at 6–10.

195. *Id.*; *see also* News Release, Am. Psychiatric Ass'n, *DSM-5* Draft Criteria Open for Public Comment (May 2, 2012), *available at* http://www.dsm5.org/Newsroom/Documents/12-24-dsm-5-public-comment-period[1].pdf.

quently published in the peer-reviewed literature.[196]

When finalized by the work group, proposed revisions were reviewed by the *DSM* Task Force (comprised of heads of each work group and an overall chair), a Scientific Review Committee (which rated the scientific strength of the proposal),[197] a Clinical and Public Health Committee (which commented on the clinical implications of the change), a committee of elected members from the American Psychiatric Association's ("APA") Assembly, and a Summit Group (comprising heads of the other review committees, the APA president and president-elect, and reviewers with forensic and nosologic expertise) meant to integrate the other levels of input and make a recommendation regarding acceptance, rejection, or modification.[198] Finally, all diagnoses were reviewed by the Board of Trustees of the APA, which made the final decision, taking into account the combined input of the other reviewers.[199] Hence, in a real sense, the *DSM-5* was subject to an exhaustive peer-review process, albeit not the typical one used by professional journals.

For what might be termed "atypical diagnoses"—that is, describing syndromes of mental disorders that do not appear in *DSM-5*—closer scrutiny will be needed. There may be valid diagnostic categories that, for one reason or another, were not included in the *DSM*, including alternative approaches to "cutting the pie" of mental disorders that cluster symptoms differently. Without the extensive peer-review process of the *DSM* to look to, courts will need to examine the basis for the evidence proffered, including whether it is supported by peer-reviewed publications, whether those studies have been replicated, the quality of the journals in which the research is reported, and the extent to which the evidence is generally accepted as valid. The same is true for behavioral syndromes with legal relevance that are not generally recognized as constituting mental disorders, such as battered woman's syndrome.[200]

Given that clinical approaches to functional assessments typically have not been standardized, data to support their accuracy will generally not be found in the peer-reviewed literature.[201] Courts generally have

196. *See, e.g.*, Regier et al., *supra* note 74.

197. *See* K. S. Kendler, *A History of the DSM-5 Scientific Review Committee*, 43 PSYCHOL. MED. 1793, 1793 (2013) ("The central role of the [Scientific Review Committee] was to provide external review for all proposals for diagnostic change in DSM-5, evaluate them on their level of empirical support using objectively structured rules of evidence agreed upon in advance and make appropriate recommendations to the leadership of the American Psychiatric Association.").

198. DSM-5, *supra* note 19, at 8–10.

199. *Id.*

200. *See generally, e.g.*, David L. Faigman, Note, *The Battered Woman Syndrome and Self-Defense: A Legal and Empirical Dissent*, 72 VA. L. REV. 619 (1986).

201. There are, however, exceptions to this generalization, though they are not very reassuring

allowed the admission of such clinical evaluations as representative of the standard of practice in the field. However, with the availability of structured instruments for a growing number of evaluations,[202] which usually provide data on validity and reliability—such data often having been published in peer-reviewed journals—we may see changes in judicial practices in the future. It is conceivable that experts who fail to use available, validated structured assessments will be asked to justify their choices and demonstrate the reliability of the procedures they have elected to follow.

b. *Diagnostic* Evidence on Peer Review and Publication

Insofar as testimony regarding the diagnosis and functional abilities of an individual will not usually be susceptible to prospective peer review, this criterion is of limited applicability to diagnostic evidence. In some sense, the introduction of multiple experts by one or both parties may mimic peer review by providing some indication that an expert's conclusions are not so idiosyncratic that they are not shared by at least some of the expert's peers. However, given the many variables that can impinge on the objectivity of an expert's conclusions—beginning with the obvious fact that experts are being paid by one of the parties and will have a much more limited role in the case if their conclusions are not favorable to that party[203]—there is a significant difference between the usual academic peer-review process and the mere concurrence of multiple expert witnesses. Nonetheless, in some large-scale litigation, teams of experts have been assembled and have met as a group to consider the evidence and their conclusions, a process that allows some degree of peer criticism to come into play.

Although true prospective peer review of expert testimony is difficult to imagine, *post hoc* peer review of expert testimony is not unheard of. More than two decades ago, the American Psychiatric Association issued a report suggesting mechanisms for peer review of expert psychiatric testimony, a suggestion that was picked up by the American Academy of Psychiatry and the Law ("AAPL"), the major U.S. professional organization for forensic psychiatrists.[204] At its annual meeting each

with regard to the reliability of clinical assessment. *See, e.g.*, Charles W. Lidz, Edward P. Mulvey & William Gardner, *The Accuracy of Predictions of Violence to Others*, 269 J. AM. MED. ASS'N 1007, 1007 (1993) (concluding that "[c]linical judgment adds to predictive accuracy"); Daniel C. Marson et al., *Consistency of Physician Judgments of Capacity to Consent in Mild Alzheimer's Disease*, 45 J. AM. GERIATRIC SOC. 453, 453 (1997) (concluding that "[s]tandardized assessment instruments and clinical training may improve the agreement and accuracy of physician competency judgments").

202. *See generally* GRISSO, *supra* note 153.

203. *See generally* Murrie et al., *supra* note 167.

204. *See generally* Am. Psychiatric Ass'n Council on Psychiatry & Law Task Force on Peer

year, an AAPL peer review committee provides confidential review of testimony submitted to it by experts themselves (transcripts or videotapes) and typically holds a public session to review testimony of an expert who has volunteered for the process. Some other medical specialties have developed comparable review mechanisms.[205] Were participation in this voluntary peer-review process to increase—and one could imagine academic departments of psychiatry and district branches of the APA offering such services—courts might look to whether an expert has participated in having his or her testimony peer reviewed as an indication of the desire of the expert to remain within acceptable professional standards.

iv. "Widespread Acceptance"

Widespread or "general" acceptance of the approach on which diagnostic or functional evaluations are based was the foundation of the *Frye* standard, which dominated considerations of admissibility for seventy years and is still the reigning standard in many jurisdictions.[206] Although *Daubert* does not require widespread acceptance, the Court noted that it could be taken into account as an index of the reliability of the proffered testimony.[207] Hence, the extent to which an expert's methods are generally accepted in the relevant professional community remains an important consideration today.[208]

a. Framework Evidence on "Widespread Acceptance"

Psychiatric diagnoses based on *DSM-5* categories would appear to have no difficulty meeting the widespread acceptance standard. The *DSM* is accepted as the standard of practice for mental health profes-

---

Review of Expert Testimony, *Peer Review of Psychiatric Expert Testimony*, 20 BULL. AM. ACAD. PSYCHIATRY & L. 343 (1992); Am. Psychiatric Ass'n Council on Psychiatry & Law Task Force on Peer Review of Expert Testimony, *American Psychiatric Association Resource Document on Peer Review of Expert Testimony*, 25 J. AM. ACAD. PSYCHIATRY & L. 359 (1997).

205. *See* Alexander R. Vaccaro, *Holding Medical Expert Witnesses Accountable*, AAOS NOW (Nov. 2007), http://www.aaos.org/news/bulletin/nov07/youraaos2.asp ("The [American Academy of Orthopaedic Surgeons] Professional Compliance Program Grievance Procedures allow a fellow or member to initiate a grievance against another fellow or member who allegedly violated the [Standards of Professionalism ("SOPs")] on Orthopaedic Expert Witness Testimony that were adopted by the fellowship, based on actions that occurred on or after April 18, 2005.").

206. *See supra* note 42. For a review of states that continue to apply the *Frye* standard, see Alice B. Lustre, *Post*-Daubert *Standards for Admissibility of Scientific and Other Expert Evidence in State Courts*, 90 A.L.R.5th 453 (2001).

207. *See* Daubert v. Merrell Dow Pharms., 509 U.S. 579, 594 (1993).

208. *See* James T. Richardson et al., *The Problems of Applying* Daubert *to Psychological Syndrome Evidence*, 79 JUDICATURE 10, 11 (1995) (In the post-*Daubert* world, "courts have at times sought refuge in *Frye*-like general acceptance principles to create order out of what appears to be a chaotic situation.").

sionals in the United States, and to a considerable extent in other countries as well.[209] For diagnoses that are not included in the *DSM*, including behavioral syndromes of legal interest, demonstration of general acceptance may be more difficult. Evidence in that regard can come from the published literature—especially if peer reviewed—major textbooks, and documents issued by professional organizations. The same is true for assessments of the process by which the evaluation was conducted, such as the areas of inquiry, the records reviewed, and the collateral informants interviewed.

Crucial to the outcome of these determinations will be the identification of the relevant professional community whose acceptance is being assessed. The debate on the existence of "repressed memories," which came to the attention of the courts in the 1990s, illustrates the problem.[210] By and large, cognitive psychologists who studied memory tended to discount the possibility that a person could repress (i.e., forget) a traumatic memory, only to have it reappear years or decades later.[211] In contrast, many clinical psychologists and psychiatrists who worked with victims of sexual abuse believed that they had seen instances of the phenomenon themselves, and that its reality could not be denied.[212] A court's choice of one community of professionals over the other as relevant to determination of widespread acceptance would result in very different outcomes. Along with the threshold problem (i.e., when is acceptance widespread enough for the court to recognize it as sufficient), this difficulty in the application of the general acceptance standard contributed to much criticism of the *Frye* approach.[213]

Functional assessments do not have a "bible" like the *DSM* to which to appeal to demonstrate their canonical status, though once again professional literature and the guidelines produced by professional organizations can be helpful in that regard. The wisdom in not relying on widespread acceptance as a mandatory criterion for admissibility is evident in the fact that many of the structured assessment instruments, which often yield more reliable and arguably more valid determinations

209. *See DSM*, AM. PSYCHIATRIC ASS'N, http://www.psychiatry.org/practice/dsm (last visited Mar. 30, 2015).

210. *See generally* Elizabeth F. Loftus, *The Repressed Memory Controversy*, 49 AM. PSYCHOL. 443–45 (1994) (discussing the phenomenon of repressed memories and critically evaluating techniques used to "recover[ ] memories").

211. *See Scientists and Practitioners Don't See Eye to Eye on Repressed Memory*, ASS'N FOR PSYCHOL. SCI. (Dec. 13, 2013), http://www.psychologicalscience.org/index.php/news/releases/scientists-and-practitioners-dont-see-eye-to-eye-on-repressed-memory.html.

212. *See* Loftus, *supra* note 210, at 444.

213. *See* Richardson et al., *supra* note 208, at 13, 15 (stating that application of the *Frye*-standard could lead to rejection of "valuable syndrome evidence" because of lack of widespread acceptance, and that would be contrary to the "spirit of *Daubert*").

of particular capacities than unstructured clinical examinations, are not widely used in the field.[214] Were widespread acceptance to be insisted upon and defined narrowly, such improvements could never be introduced into forensic assessment. Hence, for many assessment instruments, other indicia of reliability and validity will take precedence, including data on testability and error rates.

b. *Diagnostic* Evidence on "Widespread Acceptance"

The most likely source of disagreement among psychiatrists regarding the application of otherwise accepted framework evidence to a particular case will concern whether the circumstances of such cases would ever permit a valid diagnostic judgment. For example, in other scientific areas, such as eyewitness identification, it is widely accepted that while framework evidence regarding factors that interfere with eyewitness accuracy are valid, it is never appropriate for an eyewitness expert to testify that a particular identification was or was not accurate.[215] The same might be true in certain areas of psychiatry. For example, courts have limited expert testimony on the battered woman syndrome and gender stereotyping to framework evidence, because these forms of psychological expertise are not accepted for diagnostic purposes.[216] Similarly, as noted previously, although it might be generally accepted that a psychiatrist can assess competency of an individual defendant, it is much less widely accepted that a psychiatrist can determine whether a particular person was competent last year or will be competent next year.[217]

In general, assuming framework evidence indicating widespread acceptance of the process of evaluation employed by the psychiatric expert, the remaining issues at the *diagnostic* level will relate to whether the expert in fact followed the process described. To the extent that the testimony in question addresses issues for which clear professional standards have not been formulated—which is not unlikely given the many permutations of behavior that psychiatric experts can be asked to address—the burden should be on the expert to explain why the approach selected is likely to result in valid and reliable results. Courts will also need to determine in each case the extent to which the technique used is likely to yield information responsive to the legal issue at hand; that is, the question of "fit."

214. *See* Krauss & Scurich, *supra* note 178, at 216–17.

215. *See supra* note 51 and accompanying text.

216. *See supra* note 139 and accompanying text.

217. *See supra* notes 153–56 and accompanying text.

v. "Same Level of Intellectual Rigor"

In *Kumho Tire Co.*, the Court added to the admissibility considerations laid out in *Daubert* and provided that an expert's testimony should "employ[ ] in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."[218] This factor alone cannot substitute for the *Daubert* guidelines, or other criteria that might assist the required assessment, for several reasons. First, the field itself might not have a reliable or valid foundation, so applying the same degree of rigor would amount to no test at all. Astrology and tea-leaf reading are obvious examples, but some practices that fall within the mental health fields have dubious bases too.[219] Second, in some instances the expert's field and the courtroom are the same. Many forensic specialties, including experts claiming the ability to identify, among other things, latent fingerprints, firearms, and bite marks, exist principally—if not exclusively—for legal use. Finally, the standards for "success" in a field might be quite different than the legal use for which the evidence is proffered. Polygraphs might be highly effective deterrents when used in national security settings or for obtaining confessions from suspects, but their power to deter or obtain confessions may be wholly unrelated to their validity. Therefore, while the "same level of intellectual rigor" factor is a necessary condition, it is not, by itself, a sufficient condition for admissibility.

a. Framework Evidence on "Intellectual Rigor"

Intellectual rigor at the framework level should be manifest in the process by which the framework evidence has been developed. The multiple layers of review in the *DSM* development process—even if it left some critics unsatisfied with the criteria in one or another category—constituted an exhaustive process of consideration. If a non-standard (i.e., non-*DSM*) diagnosis is offered, the expert should be able to demonstrate that the process of validating the condition and establishing the reliability of its assessment lives up to similar professional standards. For functional assessments, if based on standardized instruments, intellectual rigor would suggest that they too be validated according to professional standards, preferably with review and publication in a peer-reviewed venue. Clinically-based evaluations of functional abilities will

218. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 152 (1999).

219. *See, e.g.*, Harrison G. Pope & James I. Hudson, *Scientific Status: Repressed Memories*, *in* MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 944, § 20 (David L. Faigman et al. eds., 2014–2015 ed.) ("[The peer-reviewed psychological] literature has failed to produce any evidence that individuals can, in fact, repress memories of traumatic events, including memories of childhood sexual abuse.").

have a more difficult time establishing their intellectual rigor, but they must reflect something more than merely the idiosyncratic approach of the evaluator.[220] Arguably, the approaches used in such assessments should be based on research data indicating their relationship to the focus of the evaluation (e.g., inquiries about substance abuse in an evaluation of violence risk are supported by a substantial literature demonstrating that substance abusers have higher rates of violent acts).

b. *Diagnostic* Evidence on "Intellectual Rigor"

Uncertainty is ubiquitous in the practice of psychiatry, other areas of medicine, and other mental health professions. As noted above, diagnoses are made on a probabilistic basis, with recognition that they may need to be adjusted over time.[221] Indeed, the practice of formulating a differential diagnosis—that is, a list of conditions that appear to be consistent with a patient's presentation, but only one of which may actually be present—as the basis for further investigation is something taught to medical students as soon as they enter the clinic. Treatments, especially in psychiatry where interventions are often aimed at symptoms rather than underlying causes, are empirical and subject to change depending on patients' responses. Estimates of functional impairment, even if accurate at the time of the evaluation, may not reflect changes in the evaluee's condition or circumstances in the future.

Physicians are taught to recognize and acknowledge this uncertainty in a variety of ways. In teaching hospitals and clinics, students and trainees present cases to groups of their peers and supervisors for discussion and consideration of the diagnostic and treatment options.[222] Differential diagnoses feature prominently in patients' charts, and some patients may be admitted to a hospital with one or more "rule-out" (commonly abbreviated "R/O") diagnoses (e.g., R/O myocardial infarction; R/O esophageal spasm), a frank acknowledgement that the actual cause of the patient's symptoms is still unclear. In discussions with patients and family members, including in the process of obtaining informed consent for recommended evaluation or treatment procedures, physi-

220. *See* Coble v. State, 330 S.W.3d 253, 279 (Tex. Crim. App. 2010) (Psychiatric expert's testimony about future dangerousness at a death penalty hearing should not have been admitted because it was based on an idiosyncratic approach, with no clear relationship to the scientific literature on prediction. Hence, "the prosecution did not satisfy its burden of showing the scientific reliability of Dr. Coons's methodology for predicting future dangerousness by clear and convincing evidence.").

221. *See supra* note 79 and accompanying text.

222. In fact, the plot of the popular television series *House* was built around these discussions. *See House*, Fox, http://web.archive.org/web/20120603155500/http://www.fox.com/house/about/ (last visited Mar. 30, 2015).

cians will often share their uncertainty, recognizing it as a piece of information that is likely to be material to decisions about patients' treatment.

When evaluations are performed for the purpose of courtroom testimony rather than treatment, the degree of uncertainty may be heightened further. Parties may be uncooperative with examinations, may have reasons for answering questions less than truthfully, or may malinger symptoms. Examinations often must be conducted in a limited period of time, without the possibility of reexamination. A full range of laboratory and x-ray tests may not be available. Opportunities for longitudinal follow-up are rare. Records may be unavailable, and collateral informants may have their own motives for obscuring the truth. Thus, one might expect that approaching these evaluations with the "same level of intellectual rigor" as in clinical practice would require affirmative acknowledgement of the uncertainty inherent in evaluation practices (i.e., in framework testimony) and in the formulation of individual diagnoses and functional assessments (i.e., in *diagnostic* testimony). This conclusion suggests that alternative interpretations of the expert's data, including other diagnostic categories or the possibility that a party has exaggerated his or her degree of functional impairment, should be considered by the expert, who should be able to articulate a reasonable basis for having rejected them.

Such acknowledgements of uncertainty, however, are less common than might be expected, in part because the contingencies of the courtroom disfavor them. Attorneys understandably want experts with opinions favorable to their case to state them firmly and without equivocation. They may encourage psychiatric experts to avoid discussions of uncertainty and may structure direct examinations to limit opportunities for such admissions. Although opposing attorneys may want to explore an expert's uncertainty during cross-examination, the dynamics of that situation often lead experts to feel that they must wholeheartedly defend the conclusions they have come to, lest any admission of uncertainty lead the entire edifice of their testimony to crumble. The common result is that conclusions on the stand are overstated because they put aside the cautions that would ordinarily be urged in the clinical setting.[223]

If courts are to apply the *Kumho Tire Co.* criterion of intellectual rigor to psychiatric testimony, they will need to examine the extent to

223. *Cf.* Paul S. Appelbaum, *Ethics and Forensic Psychiatry: Translating Principles into Practice*, 36 J. AM. ACAD. PSYCHIATRY & L. 195, 197–98 (2008) (stating that psychiatrists can "increase the reliability, validity, and probative value of [their] determinations" through structured assessments, but "acknowledging the current tendency of some forensic experts grossly to overstate" their conclusions).

which the legitimate degrees of uncertainty in approaches to, and results of, evaluations are acknowledged. The common requirement that physicians state their conclusions "to a *reasonable* degree of medical certainty" should be helpful in underscoring that absolute certainty is neither required nor likely to be obtained in most cases.[224] In essence, the requirement that psychiatric experts think about their data as critically as they do in clinical settings is the flip-side of *Daubert's* focus on the testability and rate of error inherent in an expert's evaluation. The many pressures militating against acknowledgement of uncertainty and risk of error suggest that courts may need to be proactive in considering this criterion for admissibility.

4. Unfair Prejudice

Under the Federal Rules of Evidence, trial courts can "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."[225] Unfair prejudice can result from testimony that is likely to lead the jury to focus on issues not central to the legal determination, or so overwhelm the remaining evidence that it results in its undue neglect.

i. Framework Evidence on Unfair Prejudice

Evidence establishing a framework for standard psychiatric diagnosis or assessment of legally relevant functional abilities is unlikely to be unfairly prejudicial, if presented clearly and accurately. However, courts may have greater concern about evidence regarding behavioral syndromes and their impact on a defendant's (or in civil cases, a plaintiff's) actions. Some purported behavioral syndromes, such as parental alienation syndrome,[226] are highly contested. Although these syndromes are unlikely to pass muster under the criteria of testability, known error rates, and widespread acceptance, an additional reason for excluding framework testimony about such conditions may be the concern that they would lead the jury to focus on matters that are so speculative as

224. *See generally* Jeff L. Lewin, *The Genesis and Evolution of Legal Uncertainty About "Reasonable Medical Certainty,"* 57 Md. L. Rev. 380 (1998).

225. Fed. R. Evid. 403.

226. For a discussion of both sides of the controversy over whether a legitimate syndrome exists in which one divorcing parent "alienates" the other parent from a child's affections, see generally William Bernet et al., *Parental Alienation*, DSM-5 *and* ICD-11, 38 Am. J. Fam. Therapy 76 (2010); M. Brianna Pepiton et al., *Is Parental Alienation Disorder a Valid Concept? Not According to Scientific Evidence. A Review of* Parental Alienation, DSM-5 *and* ICD-11 *by William Bernet*, 21 J. Child Sexual Abuse 244 (2012).

not to warrant their consideration.[227] Even more commonly accepted behavioral syndromes, such as battered woman's syndrome or child sex abuse accommodation syndrome, remain controversial, both as to the validity of the categories themselves and regarding the conclusions that can legitimately be drawn from them.[228] On the other hand, framework evidence regarding approaches to assessing or understanding mental states or behaviors that are accepted as sufficiently valid and reliable ought not to be excluded as unfairly prejudicial.

ii. *Diagnostic* Evidence on Unfair Prejudice

Prejudice at the level of testimony regarding diagnosis or functional impairment could result from evidence that is likely to be given undue weight by the jury. Psychiatric testimony has been a particular focus of concern in this regard, as manifest in state rules excluding psychiatric testimony on a defendant's mental state in criminal trials, except in support of an insanity defense.[229] This practice, upheld against constitutional challenge by the U.S. Supreme Court in *Clark v. Arizona*, appears to be based on a fear that such testimony will overwhelm and distort the jury's deliberations, leading them to ignore other considerations or draw unwarranted inferences regarding the ultimate issue in the case. Similar concerns have been raised regarding testimony about brain neuroimaging studies, with courts worrying that the presentation of such images (including structural and functional magnetic resonance imaging ("MRI") and positron emission tomography ("PET")), with their striking color patterns and aura of cutting-edge science, will lead jurors to confer undue authority on the conclusions of the expert.[230] In response, trial courts have sometimes prevented experts from showing the images to the jury, but allowed them to testify about their findings, hoping thereby to mitigate whatever prejudicial effect may result.[231]

The concern that psychiatric evidence may be unfairly prejudicial itself may reflect a different sort of prejudice. In particular, critics complain that psychiatric disorders are not "real" in the same way that physical illnesses are, and hence that testimony about them will only confuse

227. *See* FED. R. EVID. 702 advisory committee's note to 2000 amendments ("The trial judge in all cases of proffered expert testimony must find that it is properly grounded, well-reasoned, and not speculative before it can be admitted.").

228. *See* Faigman, *supra* note 200 and accompanying text.

229. Clark v. Arizona, 548 U.S. 735, 756–57 (2006) (rejecting defendant's claim that an Arizona law deprived him of due process when it prohibited expert testimony regarding the defendant's "mental disease or defect" unless it had "bearing on an insanity defense").

230. *See generally* Virginia Hughes, *Science in Court: Head Case*, 464 NATURE 340 (2010) (discussing use of MRI evidence in a death penalty case).

231. *Id.* at 341.

jurors.[232] Moreover, expressions of such concern often ignore data suggesting that jurors do *not* give undue weight to psychiatric testimony or to evidence based on neuroimaging and behavioral genetic findings.[233] Inherent in excluding testimony on the basis of likely prejudicial impact is the belief that cross-examination will be ineffective in placing the testimony into an appropriate context. That would appear to be an empirical question, data regarding which could improve the validity of judicial determinations of admissibility.

C. *G2i and the Evolution of Psychiatric Expert Testimony*

Like other areas of medicine and science more broadly, approaches to psychiatric assessment evolve over time as new concepts take hold and new techniques become available. Here we highlight two areas in which evolving concepts, in some cases based on new technologies, may impact psychiatric expert testimony, and briefly consider how they are likely to affect the G2i considerations discussed above.

1. Dimensional Approaches to Diagnosis

Psychiatric diagnosis, like medical diagnosis in general, traditionally has been based on dichotomous determinations of the presence or absence of a disorder. As a concrete example, either a patient meets the diagnostic criteria for major depressive disorder or he does not. The diagnostic criteria in the *DSM* generally allow for no intermediate state; like pregnancy, one cannot have a little bit of major depression.[234] This categorical approach to diagnosis meets a number of needs. For the clinician, it indicates the presence of a disorder that requires treatment and often suggests the range of options that are most likely to be effective. For the health insurer, it signals that payment for treatment is likely to fall within the scope of an insured's policy. For the legal system, it specifies that the predicate requirement for many legal determinations has been met (e.g., disability, civil commitment, insanity defense). In many practical respects, then, it is an ideal approach.

Categorical diagnoses, however, have been subject to growing criticism in recent years on the grounds that they fail to reflect the complex-

232. *See generally* Steven K. Erickson, *The Myth of Mental Disorder: Transsubstantive Behavior and Taxometric Psychiatry*, 41 Akron L. Rev. 67 (2008).

233. *See* Paul S. Appelbaum, *The Double Helix Takes the Witness Stand: Behavioral and Neuropsychiatric Genetics in Court*, 82 Neuron 946, 948 (2014); *see also* Paul S. Appelbaum & Nicholas Scurich, *Impact of Behavioral Genetic Evidence on the Adjudication of Criminal Behavior*, 42 J. Am. Acad. Psychiatry & L. 91 (2014).

234. However, *DSM-5* does provide for a "course specifier" of "in partial remission," indicating that full criteria for major depression have been met in the past but are no longer met. DSM-5, *supra* note 19, at 162.

ity of mental disorders. Symptoms of depression, as many people can testify from personal experience, are not just present or absent, but may be present to varying degrees. By imposing particular (critics might say arbitrary) cut-offs to determine when a diagnosis can be made (e.g., someone who meets four symptom criteria does not qualify for a diagnosis of major depressive disorder, whereas someone with five symptoms does), contemporary psychiatric diagnosis often fails to reflect either the experience of patients or the biology of the disorders. Hence, there has been considerable discussion in the psychiatric and psychological literatures about the possibility of moving toward a dimensional approach to diagnosis.[235] Under a dimensional approach, rather than determining whether someone does or does not meet the criteria for depression, a clinician would assess the extent to which a person manifests symptoms of depression along a continuum of severity.

When the *DSM-5* revision process began, its leaders announced that dimensional approaches would feature prominently in the new edition.[236] In the end, however, *DSM-5* retained an almost entirely categorical approach.[237] As a field, psychiatry was not yet ready to give up its discrete diagnostic categories and confront the clinical and practical consequences of a dimensional approach. But the pressure to move toward dimensional diagnoses will continue, raising the question of the implications of such a move for psychiatric expert testimony.[238] Arguably, dimensional characterizations of parties in civil or criminal cases could yield more valid evidence at both the framework and *diagnostic* levels. To the extent that structured approaches to dimensional assessment are adopted, diagnostic conclusions may be more reliable as well.

---

235. *See generally* Darrel A. Regier, *Dimensional Approaches to Psychiatric Classification: Refining the Research Agenda for DSM-V: An Introduction*, 16 INT. J. METHODS PSYCHIATRIC RES. S1 (2007), *available at* http://www.dsm5.org/Research/Documents/Regier_Dimensional%20Approaches_An%20Introduction.pdf.

236. *See generally DSM-5's Integrated Approach to Diagnosis and Classifications*, AM. PSYCHIATRIC ASS'N 1 (2013), *available at* http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCQQFjAAahUKEwit8Lfj1Y_HAhVC8oAKHQ9tDOI&url=http%3A%2F%2Fwww.psychiatry.org%2FFile%2520Library%2FPractice%2FDSM%2FDSM-5%2FDSM-5-Integrated-Approach.pdf&ei=btDAVa2BLsLkgwSP2rGQDg&usg=AFQjCNF7UhrFMLsSsV1IkPUDBgIBCOUxeA&sig2=7vekIyol6xsw7gDRZiWBmQ&bvm=bv.99261572,d.eXY.

237. One exception is the new diagnostic category of autism spectrum disorder, which recognizes a gradient of severity of symptoms in patients who would previously have been diagnosed with Asperger's syndrome, autism, or pervasive developmental disorder.

238. The National Institute of Mental Health is promoting a still more radical dimensional approach to psychiatric diagnosis: Research Domain Criteria ("RDoC"). RDoC abandons traditional symptom clusters altogether, rating patients along dimensions of biological function (e.g., cognitive control, perception, approach motivation). *See Research Domain Criteria (RDoC)*, NAT'L. INST. MENTAL HEALTH, http://www.nimh.nih.gov/research-priorities/rdoc/index.shtml (last visited Mar. 30, 2015). Although beginning to be adopted for research purposes, RDoC is not yet validated and not intended at this point for use in the clinical realm.

However, dimensional diagnoses may raise concerns regarding legal fit and helpfulness. To the extent that the law continues to think categorically, requiring dichotomous choices to be made about the presence or absence of a disorder as a predicate to a variety of legal claims, dimensional assessments may neither meet the law's need for a threshold criterion for the application of certain legal rules nor be helpful to factfinders trying to apply those rules. Of course, there may be ways of adapting legal criteria over time to a new, dimensional diagnostic framework, such as identifying cut-off points on the spectrum of disorder that would satisfy the mental disorder predicate. But the prospect of having dimensional constructs replace categorical diagnoses illustrates that improvements in diagnostic validity may not necessarily enhance the overall value of diagnostic assessments in court.

2. Neuroscience-Based Testimony

Whereas dimensional approaches to diagnosis have not yet made their way into the courts, testimony supported by neuroimaging and genetic findings is becoming increasingly common.[239] At this point, psychiatric diagnosis remains almost entirely a clinical enterprise based on the signs and symptoms of mental disorders. However, there is a growing body of research demonstrating changes in brain structure and function associated with various psychiatric disorders, which are detectable with brain imaging technologies.[240] Likewise, active exploration is underway of the genetic correlates of mental disorders.[241] Although many of the major psychiatric disorders appear to have complex genetic roots that may not be fully explicated for many years, genes that increase the risk of such conditions have been identified and are being introduced into evidence to support the validity of clinical diagnoses.[242] And there are rarer conditions for which genetic testing is the definitive determinant of diagnosis (e.g., fragile-X syndrome,[243] DiGeorge syndrome[244]). Should research on the biological correlates of mental disorders develop to the point where neuroimaging and genetic data can

239. *See* Appelbaum, *supra* note 233, at 946.

240. *See* Martha J. Farah & Seth J. Gillihan, *Diagnostic Brain Imaging in Psychiatry: Current Uses and Future Prospects*, 14 Virtual Mentor 464, 466 (2012).

241. *See generally* Kenneth S. Kendler, *What Psychiatric Genetics Has Taught Us About the Nature of Psychiatric Illness and What Is Left to Learn*, 18 Molecular Psychiatry 1058 (2013).

242. *See generally* Deborah W. Denno, *Courts' Increasing Consideration of Behavioral Genetics Evidence in Criminal Cases: Results of a Longitudinal Study*, 2011 Mich. St. L. Rev. 967 (2011).

243. *See generally* Kathryn B. Garber, Jeannie Visootsak & Stephen T. Warren, *Fragile X Syndrome*, 16 Eur. J. Hum. Genetics 666 (2008).

244. *See generally* P.J. Scambler et al., *Velo-Cardio-Facial Syndrome Associated with Chromosome 22 Deletions Encompassing the DiGeorge Locus*, 339 Lancet 1138 (1992).

accurately identify the presence or absence of a disorder, concerns about the validity and reliability of many psychiatric diagnoses may diminish considerably.

In addition to their use for diagnostic purposes, neuroscience-based techniques are being applied to functional assessments. Functional MRI ("fMRI"), which can detect changes in brain activity, is being introduced in capital sentencing hearings to suggest that impaired brain functioning may reduce a defendant's ability to control his behavior and thus render a death penalty less appropriate.[245] Behavioral genetic evidence regarding the presence of gene variants that are associated with increased impulsivity and higher rates of criminal behavior often accompanies neuroimaging evidence.[246] Quantitative electroencephalography ("QEEG"), which measures brainwave activity, has been introduced for similar purposes.[247] Other uses for neuroimaging with which the courts must contend include the evaluation of pain, determination of brain trauma, and assessment of the veracity of witnesses (although efforts to admit evidence on EEG or fMRI-based lie detection have generally been unsuccessful).[248] Active research is exploring the use of structural MRI for the prediction of criminal recidivism,[249] and beginning to identify the correlates of the intentional states to which the law has looked to allocate responsibility for the consequences of behavior (e.g., negligent, reckless, and purposeful states of mind). It seems likely that this list will expand considerably as efforts are made to apply neuroimaging to a wider range of functional attributes.

Optimists suggest that the introduction of neuroscience-based data will improve the reliability and validity of expert testimony. However, a copious literature testifies to the concerns that these developments have raised.[250] Courts faced with determinations regarding the admissibility

245. *See* Abram S. Barth, *A Double-Edged Sword: The Role of Neuroimaging in Federal Capital Sentencing*, 33 AM. J.L. & MED. 501, 521 (2007).

246. *See* Appelbaum, *supra* note 233, at 946.

247. *See* Gerald Gluck, *QEEG Accepted in Death Penalty Trial in* Florida v. Nelson, 39 BIOFEEDBACK 74, 74–77 (2011).

248. *See* Michael E. Robinson, Roland Staud & Donald D. Price, *Pain Measurement and Brain Activity: Will Neuroimages Replace Pain Ratings?*, 14 J. PAIN 323 (2013); Leonard Berlin, *Neuroimaging, Expert Witnesses, and Ethics: Convergence and Conflict in the Courtroom*, 5 AJOB NEUROSCI. 3 (2014); Daniel D. Langleben & Jane Campbell Moriarty, *Using Brain Imaging for Lie Detection: Where Science, Law, and Policy Collide*, 19 PSYCHOL. PUB. POL'Y & L. 222, 225 (2013).

249. *See* Eyal Aharoni et al., *Neuroprediction of Future Rearrest*, 110 PROC. NAT'L ACAD. SCI. 6223 (2013); Eyal Aharoni et al., *Predictive Accuracy in the Neuroprediction of Rearrest*, 9 SOC. NEUROSCI. 332 (2014).

250. *See, e.g.*, C.C. Meltzer et al., *Guidelines for the Ethical Use of Neuroimages in Medical Testimony: Report of a Multidisciplinary Consensus Conference*, 35 AM. J. NEURORADIOLOGY 632, 632–35 (2014).

of neuroimaging and other data will need to apply the full range of criteria identified by Faigman, Monahan, and Slobogin,[251] but this time to a complex, unfamiliar, and rapidly developing area of science. A new cast of experts—including neurologists, neuroradiologists, neuroimagers, and geneticists—will be taking the witness stand together with or in place of the psychiatrists and psychologists who are familiar visitors to the courts, each with their own specialized language and approaches to assessment. Perhaps most difficult to resolve in many cases will be the question of legal fit, that is, whether evidence of genetic predispositions or atypical patterns of brain activation actually speak to the issues of concern to the law, including responsibility and punishment.[252] These developments are only likely to increase the relevance of G2i analysis in criminal and civil cases alike.

V. Conclusion

The ubiquity of G2i issues in psychiatric expert testimony suggests the importance of courts addressing them in a systematic way. When psychiatric experts are called to the stand, it is generally to provide *diagnostic* evidence regarding an individual plaintiff or defendant. However, with uncommon exceptions, that evidence is based on an underlying framework of diagnostic or functional assessment with greater or lesser validity. And the application of such a framework to the case in question, translating group findings to individual conclusions, will reflect varying degrees of reliability and validity. Once this two-stage process is recognized for what it is, the door is open for judges to undertake a systematic examination of G2i concerns that should be addressed in determining the admissibility of the evidence.

The burden of providing the information necessary for this examination, of course, falls on the party propounding the testimony. All too often today, psychiatric expert witness testimony is accepted after a perfunctory process of establishing the clinical credentials of the expert, to the exclusion of careful consideration of the basis for the expert's testimony. In keeping with established standards for admissibility of expert evidence, however, courts should insist on more. Experts should be responsible for describing the conceptual framework on which their testimony is based, along with the data supporting that framework. In some cases, this process will be facilitated by reliance on bodies of data that have been subject to extensive vetting, such as with regard to *DSM* diagnoses. In other cases, such as when diagnostic categories that are not in

251. *See* Faigman, Monahan & Slobogin, *supra* note 3, at 440.

252. *See* Stephen J. Morse, *Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic Note*, 3 Ohio St. J. Crim. L. 397, 400–01 (2006).

the standard nosology are being used (e.g., parental alienation syndrome), *de novo* examination of the relevant data may be required. Along with consideration of relevance, helpfulness, and prejudice, careful examination of the reliability and validity of the principles and procedures underlying framework evidence is essential.

Even valid frameworks, however, do not necessarily yield accurate conclusions regarding specific persons. Membership in a group (e.g., persons with post-traumatic stress disorder ("PTSD")) does not imply that an individual necessarily shares the average tendencies of the group (e.g., the fact that most people with PTSD startle easily does not mean that a particular defendant with PTSD startles easily). The uncertainties involved in inferring individual characteristics from group data imply that in a given case conclusions based on group membership can be stated with greater or lesser degrees of certainty, but never without residual doubt. Here again, experts should be called on to identify the basis for their inferences, the data supporting their approach, and the degree of confidence that reasonably can be placed in their judgments. Even if deemed admissible, the weight given to expert testimony may also be affected by G2i concerns, and juries may need to be instructed accordingly.

This process of taking G2i considerations into account in determining admissibility is clearly more demanding of judicial time and attention than much current practice. Although it should improve the quality of psychiatric expert testimony as a whole, it will also result in the exclusion of testimony that courts may once have relied on routinely to help resolve difficult disputes.[253] As G2i issues are recognized more broadly in other areas of expert testimony, similar effects are likely across the board. Whatever the costs, however, the G2i problem cannot be ignored without undermining the conceptual basis for expert testimony as a whole and, hence, is in urgent need of judicial attention.

253. As an example, psychiatric and psychological testimony in child custody disputes about which disposition is in the best interests of the child—much of which rests on poor or nonexistent empirical frameworks—would be at high risk for exclusion under a more rigorous approach, leaving courts to resolve these contentious cases on some other grounds. *See* Scott & Emery, *supra* note 104, at 71.