

1-1-2003

Danger at the Edge of Chaos: Predicting Violent Behavior in a Post-Daubert World

Erica Beecher-Monas
Wayne State University

Edgar Garcia-Ril
University of Arkansas for Medical Sciences

Recommended Citation

Beecher-Monas, Erica & Garcia-Ril, Edgar. Danger at the Edge of Chaos: Predicting Violent Behavior in a Post-Daubert World. 24 *Cardozo L. Rev.* 1845 (2003)
Available at: <https://digitalcommons.wayne.edu/lawfrp/152>

This Article is brought to you for free and open access by the Law School at DigitalCommons@WayneState. It has been accepted for inclusion in Law Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.

DANGER AT THE EDGE OF CHAOS: PREDICTING VIOLENT BEHAVIOR IN A POST-*DAUBERT* WORLD

*Erica Beecher-Monas**
*Edgar Garcia-Rill, Ph.D.***

INTRODUCTION

For nearly twenty years we have known that psychiatrists cannot predict whether a person who has committed a violent act will be violent in the future.¹ Neither can lay people.² The very best anyone can do is speculate. Even the most scientific predictions based on thorough examination, diagnosis of mental symptoms, past patterns of behavior, and probabilistic assessment are wrong nearly as often as they are right.³ The most common

* Associate Professor of Law, University of Arkansas at Little Rock William H. Bowen School of Law; J.S.D. Columbia University School of Law, J.D. University of Miami School of Law; M.S. (Anatomy) University of Miami. Many thanks to J. Thomas Sullivan. Thanks also to Amy Dunn for her able and enthusiastic research assistance.

** Director, Center for Translational Neuroscience, Professor of Anatomy and Neurobiology University of Arkansas for Medical Sciences, Ph.D. McGill University (Montreal), postdoctoral training U.C.L.A.

¹ See generally CHRISTOPHER WEBSTER ET AL., *THE VIOLENCE PREDICTION SCHEME: ASSESSING DANGEROUSNESS IN HIGH RISK MEN* (1994) (detailing studies demonstrating the inaccuracy of violence risk predictions).

² See VERNON L. QUINSEY ET AL., *VIOLENT OFFENDERS: APPRAISING AND MANAGING RISK* 62 (1998) (noting that “laypersons and the clinicians had few differences of opinion” about assessments of dangerousness, and that neither had much accuracy).

³ See, e.g., Charles W. Lidz et al., *The Accuracy of Predictions of Violence to Others*, 269 JAMA 1007, 1010 (1993) (concluding that “clinicians are relatively inaccurate predictors of violence”). In this study, when clinicians divided institutionalized men into two groups, “violent” and “nonviolent,” and examined their behavior more than three years later, 53% of the “violent” group had committed acts of violence, as opposed to 36% of the “nonviolent” group. See *id.* at 1007. Random predictions would have a sensitivity and specificity of 50%. See *id.* at 1009. Thus, while the results are better than chance, the low sensitivity and specificity of the predictions show “substantial room for improvement.” *Id.* Sensitivity is the percentage of times that a test correctly gives a positive result when the individual tested actually has the characteristic in question. See Bruce R. Parker & Anthony F. Vittoria, *Debunking Junk Science: Techniques for Effective Use of Biostatistics*, 66 DEF. COUNS. J. 33, 34 (1999). Specificity is the percentage of times a test

courtroom predictions—frequently based solely on hypotheticals—are wrong twice as often as they are right.⁴ Nevertheless, legislatures and courts continue to demand expert risk assessments.⁵ Experts routinely testify in courtrooms across the country that their scientific expertise gives them enough insight to be able to predict “with reasonable medical certainty” that a capital defendant will commit future acts of violence, and therefore merits the death penalty.⁶

These experts are allowed to testify because, as a constitutional matter, all that is required is a floor of reliability and relevance.⁷ Evidentiary rules do not generally apply at sentencing hearings.⁸ The floor of relevance and reliability in the context of

correctly reports that a person does not have the characteristic under investigation. *See id.* Actuarial studies, though more accurate than clinical predictions, still predict with less than stellar accuracy: when scores on the most accurate of the actuarial instruments, the VRAG, “were dichotomized into ‘high’ and ‘low’ risk groups, the results indicated that 55% of the ‘high scoring’ subjects committed violent recidivism, compared with 19% of the ‘low scoring’ group.” John Monahan, *Violence Risk Assessment: Scientific Validity and Evidentiary Admissibility*, 57 WASH. & LEE L. REV. 901, 908 (2000); *see also infra* Part IV.B.2.c (discussing the VRAG).

⁴ *See* JOHN MONAHAN, PREDICTING VIOLENT BEHAVIOR: AN ASSESSMENT OF CLINICAL TECHNIQUES 1 (1981) (surveying the major studies of clinical prediction of future dangerousness and finding that psychiatrists had about a one in three chance of predicting future dangerousness correctly).

⁵ *See* WEBSTER ET AL., *supra* note 1, at 17 (decrying the legal and political pressures on experts to make violence risk assessments in the face of their inherent inaccuracy).

⁶ Future dangerousness predictions are also used in psychiatric commitment proceedings, parol determinations, sexual predator determinations, bail decisions, and a variety of other legal settings. *See id.* at 17-21 & n.1 (noting the variety of settings in which future dangerousness predictions are demanded of mental health experts and observing that “[c]ourts and legislatures impute powers to psychiatry and the related professions which by and large they do not possess”). Because “death is different” and the stakes are considerably higher, this Article limits its discussion to the context of capital sentencing testimony, although the scientific validity analysis is equally applicable to these other settings.

⁷ Even the American Psychiatric Association acknowledged that the unreliability of clinical predictions of dangerousness is an “established fact.” *Barefoot v. Estelle*, 463 U.S. 880, 920 (1983). The wholesale professional condemnation of the reliability of this type of testimony failed to move a majority of the Supreme Court, however, either in *Barefoot* or in subsequent cases.

⁸ *See, e.g.*, ALA. CODE § 13A-5-45 (Repl. 1994) (providing that “[a]ny evidence which has probative value and is relevant to sentence shall be received at the sentence hearing regardless of its admissibility under the exclusionary rules of evidence”); U.S. SENTENCING GUIDELINES MANUAL § 6A1.3(a) (2001) (“In resolving any dispute concerning a factor important to the sentencing determination, the court may consider relevant information without regard to its admissibility under the rules of evidence applicable at trial . . .”). Hearsay, for example, is frequently admissible in sentencing hearings. *See, e.g.*, *Williams v. New York*, 337 U.S. 241, 151 (1949); *Todd v. Schomig*, No. 98-428-GPM, 2002 WL 392988, at *8 (7th Cir. Mar. 14, 2002). So is testimony about unadjudicated prior offenses. *See, e.g.*, *United States v. Lee*, 274 F.3d 485 (8th Cir. 2001); *Gilbert v. State*, 951 P.2d 98 (Okla. Crim. App. 1997). Nonetheless, both hearsay and unadjudicated offenses must have “indicia of reliability” and relevance to be admissible even at sentencing hearings. *See Dawson v. Delaware*, 503 U.S. 159 (1992) (finding

science however, is that the testimony reflect valid science. What goes into assessing whether testimony reflects valid science has been addressed in the context of evidentiary rules by the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,⁹ and its progeny. Although those decisions are evidentiary rather than constitutional, the floor of relevance and reliability is of even greater importance in the constitutional setting. Thus, the *Daubert* requirement of scientific validity is based on sound logic and ought to be applied to the constitutional issue of whether the expert testimony makes the trial so unfair that it violates due process.¹⁰

An astrologer, for example, would not be permitted to testify about future dangerousness, either as a constitutional matter or under rules of evidence, because the testimony would be misleading. Actually, it would be no more misleading than permitting medical experts to testify about future danger, since lay people can predict future dangerousness as well as medical experts.¹¹ A trial fraught with error espoused by experts can hardly be fair and certainly would not be helpful to the jury.¹² Although the testimony of clinicians about future dangerousness offers little more than that of an astrologer, such clinical testimony is pervasive, and courts persist in circumventing any inquiry into the scientific validity of expert future dangerousness predictions. This is an important concern because giving the imprimatur of science to chicanery undermines our justice system. Because of this concern, researchers undertook a series of studies in order to improve the accuracy of such predictions.

Nonetheless, courts persist in circumventing any inquiry into the scientific validity of expert future dangerousness predictions. This is a concern because when a court permits the imprimatur of

evidence presented at sentencing hearing that convicted murderer was a member of the Aryan Brotherhood irrelevant); *United States v. Huckins*, 53 F.3d 276 (9th Cir. 1995) (to be admissible at sentencing, hearsay must have some other corroboration).

⁹ 509 U.S. 579 (1993). *Daubert* requires that expert testimony in federal courts meet minimum standards of reliability. *Id.*

¹⁰ *See, e.g., Brady v. Maryland*, 373 U.S. 83, 87 (1963) (requiring the state to provide the defense with any exculpatory evidence it possesses on reliability grounds, "not [as] punishment of society for misdeeds of a prosecutor but [as] avoidance of an unfair trial to the accused").

¹¹ *See* QUINSEY, *supra* note 2, at 62 (noting that "laypersons and the clinicians had few differences of opinion" about assessments of dangerousness). Indeed, the expert future dangerousness testimony in *Barefoot* was found to be reliable because lay witnesses were permitted to testify as to future dangerousness and lay witnesses could do no better. *See Barefoot*, 463 U.S. at 896.

¹² *See* 1 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 299 (1997) (arguing that although *Daubert* sets an evidentiary standard, and *Barefoot* sets a constitutional standard, the reliability focus of *Daubert* should inform the constitutional minimum of *Barefoot*).

science to be given to chicanery, justice becomes a mockery. Because of this concern, a series of studies was undertaken by researchers in order to improve the accuracy of such predictions.¹³ Their goal was to develop an empirically based actuarial instrument that would reflect a state of the art understanding of the factors correlated with violence and their inter-relationships.¹⁴ This Article will examine the scientific validity of this approach.

This Article proceeds in five parts. Part I will discuss the use of future dangerousness evidence in capital sentencing hearings and the applicable constitutional standard articulated in the Supreme Court's decision in *Barefoot v. Estelle*.¹⁵ Part II will outline the current standards for admissibility of expert testimony in the context of capital sentencing hearings, address the tension between the Supreme Court's *Daubert* trilogy of cases¹⁶ and *Barefoot*, and discuss the currents of unease beginning to stir the courts, concluding that the reliability standards of *Daubert* should indeed apply to expert testimony on future dangerousness.

In Part III, this Article will explore the latest scientific research about brain structure and function, and its implications for predicting future dangerousness, and make some suggestions about how this research could be used to aid the factfinder faced with having to make a life or death decision. Part III will also address risk analysis as a new paradigm for assessing future dangerousness and discuss the development and use of actuarial instruments in making such assessments. In Part IV, this Article will analyze the validity of actuarial instruments currently used to assess the risk of future violence, and will explore whether actuarial-based expert testimony can aid a jury's sentencing decision. Part V discusses the prediction of dangerousness in light of cognitive psychology, suggesting that testimony about actuarial instruments may help the jury resolve a key question. This Article concludes that although actuarial instruments should be used with caution, they offer improvements over the unaided judgment of juries and over the kind of unscientific assertions about future

¹³ See WEBSTER ET AL., *supra* note 1, at xi, xii (describing the genesis and goals of the violence risk assessment guide).

¹⁴ See QUINSEY, *supra* note 2, at 190; WEBSTER ET AL., *supra* note 1, at xii.

¹⁵ 463 U.S. 880 (1983) (permitting experts to testify about future dangerousness as a constitutional matter).

¹⁶ See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999) (extending the scope of the *Daubert* inquiry to technical as well as scientific evidence); *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997) (reiterating the trial judge's mandate to review testimony for scientific validity and "fit"); *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993) (instructing the federal judiciary to make admissibility determinations based on analyzing the scientific validity of the proffered testimony, and on whether the testimony "fits" the issues in the case).

dangerousness currently typical in capital sentencing proceedings.

I. FUTURE DANGEROUSNESS TESTIMONY IN THE COURTS

A. *State Courts and Future Dangerousness Evidence*

In capital sentencing hearings, future dangerousness may become an issue in a number of ways. It may be introduced as an aggravating circumstance, either statutory¹⁷ or non-statutory.¹⁸ It may be introduced by the prosecution through proffered expert testimony, cross-examination of a defense expert, or through the prosecutor's argument.¹⁹ Because the capital sentencing statutes of some states list absence of future dangerousness as a mitigating circumstance,²⁰ the defense may also introduce future dangerousness testimony through an expert, through character witnesses, or simply through argument.²¹

A number of states, such as California and Florida, have common law prohibitions against the prosecution introducing expert testimony on the issue of future dangerousness.²² This can put the defense between a rock and a hard place however, because if the defense offers any expert testimony regarding mental health (either as part of an insanity defense or as a mitigating factor at sentencing), that may be seen as "opening the door" to future dangerousness testimony by prosecution experts.²³ For example, in

¹⁷ Future dangerousness is a statutory aggravating factor in six states: Idaho, Oklahoma, Oregon, Texas, Virginia, and Wyoming. See IDAHO CODE § 19-2515(h) (Michie Supp. 2001); OKLA. STAT. ANN. tit. 21, § 701.12 (West 1983); OR. REV. STAT. § 163.150(b) (1999); TEX. CRIM. PROC. CODE ANN. art. 37.071, § 2(b) (West Supp. 2001); VA. CODE ANN. § 19.2-264.2 (LEXIS Repl. 2000); WYO. STAT. ANN. § 6-2-102(e) (West 2001).

¹⁸ See, e.g., *People v. Evans*, 708 N.E.2d 1158 (Ill. 1999); *State v. Hughes*, 521 S.E.2d 500 (S.C. 1999).

¹⁹ See, e.g., *Nethery v. State*, 692 S.W.2d 686, 708 (Tex. Crim. App. 1985) (upholding admissibility of expert testimony as to defendant's future dangerousness).

²⁰ COLO. REV. STAT. § 16-11-103(4)(k) (2000); IND. CODE ANN. § 35-38-1-7.1(c)(8) (LEXIS Supp. 2001); MD. ANN. CODE art. 27, § 413(g) (Supp. 2000).

²¹ See, e.g., *State v. Davis*, 477 A.2d 611 (N.J. 1984) (per curiam) (upholding trial court's admission of sociologist's expert opinion on defendant's potential for rehabilitation).

²² See *People v. Rodrigues*, 885 P.2d 1, 64 (Cal. 1994); *Haliburton v. Dep't Corr.*, 160 F. Supp. 2d 1382 (S.D. Fla. 2001).

²³ See *Ex parte Brewer*, 50 S.W.3d 492, 494 (Tex. Crim. App. 2001). The court stated that:

Presenting expert psychiatric and psychological evidence could be a 'double edged sword' in that negative information about Applicant could be exposed and exploited by the prosecution. Possibly the jury would consider such evidence adversely to Applicant in answering the 'future danger' punishment

Ruiz v. Norris,²⁴ although future dangerousness is not an aggravating factor under Arkansas law, an Arkansas federal district court held that prosecution expert testimony concerning the defendants' future dangerousness was admissible because the defendants had proffered their own expert to testify that they had "mellowed" with age.²⁵ Therefore, even where it is not a statutory aggravating factor, future dangerousness testimony has a way of creeping into evidence.

Even without expert predictions, the prosecution may attempt to demonstrate future dangerousness by showing that the defendant has a past history of violent crime.²⁶ Although a pattern of previous violence is one of the strongest predictors of future violence, the problem with sentencing a defendant based on unadjudicated conduct is that the evidence of such conduct is often flimsy at best. Indeed, some courts are beginning to show concern about the reliability issues implicated in admitting evidence of unadjudicated crimes. For example, in considering the constitutional implications of admitting evidence of prior unadjudicated offenses in order to prove future dangerousness, the Eastern District of Virginia acknowledged a heightened reliability standard for capital sentencing and recognized that, although it is not bound to observe the rules of evidence regarding admissibility, the court should not permit "an evidentiary free-for-all that undermines reliability."²⁷ As a result, the court required a hearing outside the presence of the jury to determine the reliability of the evidence of the unadjudicated prior offenses. A similar approach to expert testimony—a *Daubert*-like hearing outside the presence of the jury to determine scientific validity of the expert testimony—is similarly necessary to meet constitutional demands for relevance and reliability of future violence predictions.

special issue.

Id.

²⁴ 868 F. Supp. 1471 (E.D. Ark. 1994).

²⁵ *Id.* at 1532. Although the court acknowledged that the prosecution's expert was not testifying in rebuttal, because the defendants' expert had not testified to their lack of dangerousness, it nonetheless felt that the defendants had opened the door to future dangerousness testimony. *See id.*

²⁶ *See, e.g., Gilliard v. Scroggy*, 847 F.2d 1141 (5th Cir. 1998) (finding that statutory aggravating factor allowing jury to consider past convictions for violent offenses "implicitly" posed future dangerousness inquiry).

²⁷ *United States v. Beckford*, 964 F. Supp. 993, 1002 (E.D. Va. 1997).

B. Barefoot and the Constitutionality of Future
Dangerousness Testimony

The Supreme Court's death penalty jurisprudence since *Furman v. Georgia*²⁸ has focused on balancing consistency in decisionmaking with individual fairness.²⁹ The Supreme Court set the minimal standard for imposing the death penalty as "the evenhanded, rational, and consistent imposition of death sentences under law."³⁰ According to the Supreme Court, the jury's "reasoned moral response" to the evidence and arguments at sentencing must be supported by information sufficient and relevant for reliable rational decisionmaking.³¹ Moreover, in *Gregg v. Georgia*,³² the Court warned that "accurate sentencing information is an indispensable prerequisite to a reasoned determination of whether a defendant shall live or die by a jury of people who may never before have made a sentencing decision."³³ Under each of these precepts, using scientifically flawed information appears to be impermissible.

Nevertheless, in *Barefoot v. Estelle*,³⁴ the Supreme Court upheld expert testimony about future dangerousness that even the majority acknowledged was completely unsupported by scientific evidence. At issue was the constitutionality of permitting psychiatrists to testify about the defendant's future behavior, given that such predictions are wrong two out of three times. Reliability, then as now, was the touchstone for admissibility of evidence at sentencing.³⁵ Before the Court was an amicus brief by the American Psychiatric Association explaining that no one (including psychiatrists) can predict with any degree of reliability that an individual will commit other crimes in the future.³⁶

At the sentencing proceeding in *Barefoot*, two psychiatrists

²⁸ 408 U.S. 238 (1972) (invalidating then-current death penalty statutes as unconstitutionally arbitrary).

²⁹ See Jonathan R. Sorensen & Rocky L. Pilgrim, *An Actuarial Risk Assessment of Violence Posed by Capital Murder Defendants*, 90 J. CRIM. L. & CRIMINOL. 1251, 1251-52 (2000).

³⁰ *Jurek v. Texas*, 428 U.S. 262, 276 (1976).

³¹ See *Penry v. Lynaugh*, 492 U.S. 302, 319 (1989).

³² 428 U.S. 153 (1976) (plurality opinion).

³³ *Id.* at 190.

³⁴ 463 U.S. 880 (1983).

³⁵ See *Flores v. Johnson*, 210 F.3d 456, 464 n.10 (5th Cir. 2000) (Garza, J., concurring) (noting that "the cardinal concern of the rules of admissibility for expert testimony—reliability . . . is also the paramount concern in addressing the constitutionality of capital sentencing procedures").

³⁶ See *Barefoot v. Estelle*, 463 U.S. 880, 899 (1983).

testified that the defendant “would probably commit further acts of violence and represent a continuing threat to society.”³⁷ They did not base their opinions on any personal examination of the defendant, nor upon any history of past violent behavior (the defendant had prior convictions for drug offenses and unlawful possession of firearms and had escaped from jail where he was being held on charges of statutory rape, but had no history of violent crime). Instead, the experts based their testimony on a hypothetical question based on the crime and the defendant’s conduct.³⁸ The Supreme Court upheld the admissibility of such testimony, remarking that disallowing it would be like “disinvent[ing] the wheel.”³⁹ Because courts had traditionally admitted such testimony, the Court refused to overturn its precedent. The Court acknowledged the American Psychiatric Association’s opposition to future dangerousness testimony due to the extreme unreliability of such evidence. Nonetheless, it found that because the Association did not claim that psychiatrists were *always* wrong with respect to future dangerousness predictions—only that they were wrong more often than not—it would not exclude such testimony, since the testimony comported with the state’s rules of evidence.⁴⁰

The issue in *Barefoot* was constitutional: Could the state sentence a defendant to death based on scientifically questionable testimony? The Supreme Court decided that it could, because the state’s evidentiary rules permitted such testimony. The Supreme Court distinguished its decision in *Barefoot* from scientific evidence cases in which testimony about future dangerousness had been disallowed by explaining that *Barefoot* sought a constitutional rule barring an entire category of expert testimony.⁴¹ The Court was “not persuaded that such testimony is almost entirely unreliable,” and so found that the adversary system would be competent to take account of its shortcomings.⁴² As a result, the Court found “no constitutional barrier to applying the ordinary

³⁷ *Id.* at 884.

³⁸ The prosecutor’s hypothetical asked the psychiatrists to assume a number of facts (taken from the testimony at trial): conviction for five nonviolent criminal offenses; arrests for sexual offenses against children; a bad reputation in the eight communities the defendant had lived in over ten years; unemployment during the two months preceding the crime; drug use; boasting of plans to commit crimes to acquaintances; shooting a police officer without provocation from a distance of six inches; and acting as though there were nothing unusual after the crime. See Brief of Amicus Curiae American Psychiatric Association at 5, *Barefoot* (No. 82-6080).

³⁹ *Barefoot*, 463 U.S. at 896.

⁴⁰ See *id.* at 904.

⁴¹ See *id.* at 899.

⁴² *Id.*

rules of evidence governing the use of expert testimony.”⁴³ The ordinary rules of evidence have changed, however,⁴⁴ and these changes have illuminated the extreme unreliability and irrelevance of expert testimony that has no empirical foundation masquerading as science.

II. *DAUBERT* AND EXPERT PREDICTIONS OF VIOLENCE

A. *The Daubert Trilogy*

The Supreme Court has altered dramatically what counts as scientific knowledge in courts of law through a trilogy of cases addressing scientific evidence. In this transformative series of cases, the Supreme Court has demanded that judges examine the empirical basis for statements made by experts in federal courts. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,⁴⁵ the Supreme Court laid the groundwork for this transformation by requiring district court judges to evaluate the scientific validity and “fit” of expert testimony. The objective of the *Daubert* analysis is to ensure the reliability and relevance of expert testimony and to be certain that an expert employs the same professional standards of intellectual rigor in the courtroom as is expected in the practice of the relevant field.⁴⁶ The standard of relevance and reliability in the context of scientific testimony now requires scientific validity.⁴⁷

In *General Electric Co. v. Joiner*,⁴⁸ the Court reiterated the *Daubert* standards, expounded on its notion of “fit,” and explained

⁴³ *Id.* at 904.

⁴⁴ *See id.* at 898 (noting that “the rules of evidence generally extant at the federal and state levels anticipate that relevant, unprivileged evidence should be admitted and its weight left to the factfinder, who would have the benefit of cross-examination and contrary evidence by the opposing party”). Post-*Daubert*, the rules of evidence require judges to act as gatekeepers to ensure that only scientifically valid expert testimony be admitted.

⁴⁵ 509 U.S. 579 (1993).

⁴⁶ *See* Paul C. Gianelli, *Daubert: Interpreting the Federal Rules of Evidence*, 15 CARDOZO L. REV. 1999, 2001-03 (1994).

⁴⁷ *See id.* at 2021 (observing that “*Daubert* required a higher standard of admissibility for money damages than *Barefoot* required for the death penalty”). Although the argument has been made that *Daubert* is not technically inconsistent with *Barefoot* because *Daubert* involved interpretation of the Federal Rules of Evidence while *Barefoot* involved interpretation of the Due Process Clause of the Constitution, nearly everyone acknowledges the tension between the two decisions. *See, e.g.*, Craig J. Albert, *Challenging Deterrence: New Insights on Capital Punishment Derived from Panel Data*, 60 U. PITT. L. REV. 321, 338 (1999) (asserting that “it goes too far to say simply that *Daubert* impliedly overruled *Barefoot*[.]” but acknowledging that “they cannot co-exist as a matter of common sense”).

⁴⁸ 522 U.S. 136 (1997).

that methodology and expert conclusions are not entirely distinct inquiries: The validity of expert conclusions depends on the soundness of the methodology. Finally, in *Kumho Tire v. Carmichael*,⁴⁹ the Court explained that not only do judges have to evaluate the scientific validity of testimony based on the traditional “hard” sciences, but that they must also evaluate the validity of expert testimony based on what are often referred to as the “soft” sciences, such as psychology.⁵⁰ Technical or other specialized knowledge must similarly meet a standard of evidentiary reliability.⁵¹ Thus, *Daubert’s* general principles apply to all expert testimony.⁵²

Although directed at federal district court judges, these decisions have affected state courts also. For example, even in states that have eschewed the *Daubert* standard in favor of the old general acceptance standard, courts are responding to the pressure to rationalize their decisions by insisting that expert testimony meet standards of scientific validity.⁵³ Thus, even where the Supreme Court’s standard is not followed—that is, in many state courts—there is an increased awareness that whatever evidence is considered should be based on a sound empirical foundation.

The rationale underlying the *Daubert* trilogy’s gatekeeping requirement is that the expert’s opinion must have a reliable basis in order to be relevant.⁵⁴ Unreliable testimony cannot assist the jury, and assisting the factfinder is the only reason for admitting expert testimony. *Daubert* set out a framework for evaluation. First, a federal trial judge faced with a proffer of scientific testimony must determine whether the expert’s testimony is “scientific knowledge” that will assist the factfinder to determine a fact in issue.⁵⁵ To qualify as scientific knowledge, the Supreme Court held, “an inference or assertion must be derived by the scientific method.”⁵⁶ Astrology is therefore out. Four “general observations” guide the inquiry into scientific validity:⁵⁷

⁴⁹ 526 U.S. 137 (1999).

⁵⁰ Engineering testimony was at issue in *Kumho Tire*. *Id.* at 137.

⁵¹ *See id.* at 141. Although the Court explicitly based its rationale on the language of Rule 702 of the Federal of Rules of Evidence, finding it makes no distinction between “scientific” knowledge and “technical” or “other specialized knowledge,” the dichotomy that many of the district courts were drawing between “soft” and “hard” science made little sense.

⁵² *See id.* at 148.

⁵³ *See, e.g.,* *Blum v. Merrell Dow Pharms., Inc.*, 705 A.2d 1314, 1323 (Pa. Sup. Ct. 1997) (reviewing experts testimony under *Frye* standard and addressing scientific validity).

⁵⁴ *See Kumho Tire Co.*, 526 U.S. at 147.

⁵⁵ *See Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590 (1993).

⁵⁶ *Id.* at 590.

⁵⁷ *Id.* at 593.

testability; peer review and publication; the existence of methodological standards, including the error rate of the methodology; and general acceptance.⁵⁸ The overall goal of these “flexible”⁵⁹ guidelines” is to evaluate expert testimony by the standards scientists themselves use to critique their work.⁶⁰

These standards have had a dramatic impact. One by one, experts are successfully challenging juror (and judge) misconceptions. Take, for example, eyewitness testimony, long believed to be the epitome of reliable evidence. It turns out that the more positive an eyewitness is about the correctness of the identification, the less likely the identification is to be correct.⁶¹ More courts are admitting such expert testimony to dispel the myth that certainty of identification is any indication of its reliability.⁶² The movement toward increased empiricism in the courts has the potential to finally dispatch outdated notions about human behavior. It also has the potential to inform and revitalize the justice portion of the criminal justice system.⁶³

Radical transformations are never painless, and this one has met with its share of resistance. Many judges continue, even after *Daubert*, to circumvent scientific analysis, relying instead on rules of thumb and other short-cuts.⁶⁴ This is especially true in criminal cases, where—with the prominent exception of DNA evidence—there has often been little if any empirical support for the assertions of experts.⁶⁵ As a result, it is common for expert

⁵⁸ See *id.* at 593-94.

⁵⁹ See *id.* at 594.

⁶⁰ See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999).

⁶¹ See BARRY SCHECK, PETER NEUFELD & JIM DWYER, *ACTUAL INNOCENCE: WHEN JUSTICE GOES WRONG AND HOW TO MAKE IT RIGHT* 58 (2001).

⁶² See, e.g., *United States v. Norwood*, 939 F. Supp. 1132 (D.N.J. 1996).

⁶³ The law's moral authority, an important means of modifying human behavior, depends on the strength of its assumptions about human nature. See Owen D. Jones, *Law and Biology: Toward an Integrated Model of Human Behavior*, 8 J. CONTEMP. LEGAL ISSUES 167, 170 (1997) (explaining that in order to successfully modify human behavior in appropriate directions, “law needs a model that integrates the social sciences (upon which it currently and heavily relies) with the life sciences” and that “the success of every legal system necessarily depends, in part, on the solidity—that is, the accuracy and the predictive power—of the behavioral model on which it rests”).

⁶⁴ See Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 72-73 (1998).

⁶⁵ For example, the F.B.I. was forced to acknowledge that the dearth of empirical support for its traditional hair analysis techniques made hair identification using such techniques untrustworthy. The F.B.I.'s “white paper commentary” on microscopic hair analysis asserts that the “microscopic characteristics of hair can assist the examiner in determining the racial origin, body area and whether disease, damage or artificial treatment is present.” Douglas Deedrick, *FBI Responds to Questions Raised About Hair Comparison Analysis*, PROSECUTOR, Nov.-Dec. 1998, at 27. Although these characteristics may be “useful,” the F.B.I. recognizes that hair comparisons do not constitute a basis for personal identification. See *id.* Microscopic hair comparisons now

testimony in a toxic tort case (where the stakes are economic) to face far greater scrutiny than testimony in a criminal case (where liberty and life itself are often at stake). Expert future dangerousness testimony, proffered in capital sentencing hearings, is frequently admitted without any scrutiny whatsoever.⁶⁶

B. *Juxtaposing Daubert and Barefoot in Future Dangerousness Testimony*

The issue of future dangerousness is unquestionably crucial. Juries focus on this issue even where no testimony about future danger has been admitted either at trial or in the sentencing phase.⁶⁷ Future dangerousness testimony is thus highly relevant to the jury's decision. But this relevance is illusory if the information presented to the jury is junk in the guise of science. The paramount constitutional concerns regarding expert testimony during sentencing are still, as they were in *Barefoot*, relevance and reliability. The standards set out in the *Daubert* trilogy, however, were not met by the *Barefoot* experts, nor by any experts offering clinical judgments about future dangerousness since.⁶⁸ The question of whether clinical predictions are falsifiable is controversial.⁶⁹ The best medical diagnosis of current condition requires personal examination and patient history, analyzed against a background of empirical data about the etiology of the disease being diagnosed.⁷⁰ Falsifiability is absent, however, where experts advance no hypothesis whatsoever, but only their personal

must be used in conjunction with mtDNA analysis. *See id.*

⁶⁶ In Texas, for example, where the state must prove beyond a reasonable doubt that "there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society," there is virtually no restriction on the scope of the expert's testimony, which frequently exceeds the bounds of scientific possibility. *See* TEX. CODE CRIM. PROC. ANN. art. 37.071, § 2 (b)(1) (Vernon 2002).

⁶⁷ *See* Peter A. Barta, *Between Death and a Hard Place: Hopkins v. Reeves and the "Stark Choice" Between Capital Conviction and Outright Acquittal*, 37 AM. CRIM. L. REV. 1429, 1466 (2000) (reporting that capital jurors in every jurisdiction focus on future dangerousness even when the issue is never raised at trial).

⁶⁸ The result has been a much more stringent admissibility standard for civil cases, where money is at stake, than in criminal cases—including death penalty cases, despite judicial rhetoric about death being different. *See* L. Timothy Perrin, *Expert Witness Testimony: Back to the Future*, 29 U. RICH. L. REV. 1389, 1423 (1995) (observing that the *Daubert* decision "implicitly recognized that trusting juries with *unreliable* expert testimony is unwise, and thus imposed a reliability requirement on scientific expert testimony that predictions of future dangerousness may well not satisfy").

⁶⁹ *See, e.g.,* Lidtz et al., *supra* note 3, at 1007 (acknowledging "little definitive knowledge about how accurately clinicians make predictions of future violence").

⁷⁰ *See* WEBSTER ET AL., *supra* note 1.

prowess.⁷¹

Future dangerousness testimony based on clinical judgment alone has been overwhelmingly castigated by the profession (and so fails peer review, publication, and the general acceptance prongs of *Daubert*). Because such predictions are wrong more often than they are right, they cannot meet the error rate inquiry.⁷² Thus, it is plain that the future dangerousness testimony in *Barefoot*, which was based neither on scientific study nor on personal medical diagnosis, and did not even purport to be based on the scientific method, cannot meet criteria for valid science.

Indeed, the basis for the Supreme Court's finding it admissible was that even a layperson could testify to future dangerousness.⁷³ Unlike the laypersons involved, however, these experts had no personal knowledge of the defendant. The *Barefoot* experts, as most such experts still do in Texas, based their testimony entirely on the defendant's conduct at trial and the facts of the crime. This was enough, according to one of the experts, to demonstrate future dangerousness with "one hundred percent and absolute" accuracy.⁷⁴ That statement alone is enough to make his testimony unreliable.⁷⁵

One of the reasons for permitting all relevant evidence into a sentencing hearing, regardless of whether it can meet the rules of evidence (providing it is sufficiently reliable), is to assist the jury in making an individualized determination of whether the death penalty is appropriate given the particular circumstances of this defendant. That is similar to the issue of "fit," which the *Daubert* court explained concerns whether otherwise valid testimony will

⁷¹ Notably, it was precisely this kind of untestable psychiatric assertion that Karl Popper was attempting to discredit in his work (cited as the basis for the scientific validity inquiry in the *Daubert* opinion). A number of scholars have attempted to distinguish between testimony based on scientific research and testimony based on clinical diagnosis. See, e.g., Daniel W. Shuman & Bruce D. Sales, *The Admissibility of Expert Testimony Based Upon Clinical Judgment and Scientific Research*, 4 PSYCHOL. PUB. POL'Y & L. 1226, 1228 (1998) (asserting that "the inherent unreliability of clinical judgment and decision making" is due to using flawed heuristics).

⁷² Although they made bald assertions that they were invariably accurate, the experts in *Barefoot* offered no substantiation for their claims.

⁷³ See *Barefoot v. Estelle*, 463 U.S. 880, 896-97 (1983).

⁷⁴ *Id.* at 919 (Blackmun, J., dissenting).

⁷⁵ No one can testify with "one hundred percent accuracy." *Id.* Nor had this expert any empirical data to support his-highly improbable-statement. Such expert hyperbole is by no means uncommon. A Texas psychiatrist, who by 1992 had participated in 144 capital cases, testified in each of them that, with medical and scientific certainty, he was sure the defendant would kill again. See Joseph T. McCann, *Standards for Expert Testimony in New York Death Penalty Cases*, N.Y. ST. B.J., Jul.-Aug. 1996, at 30, 31 (outlining the prevalence of improper assessments of future dangerousness). One of those condemned was later found to be innocent of the crime. See *id.* at 32.

actually assist the factfinder.⁷⁶ The Court pointed out in *Joiner* that conclusions and methodology must have a valid connection between them.⁷⁷ Thus, unless an expert can demonstrate sound methodology and scientific reasoning, no opinion testimony is admissible.

If testimony is without foundation—that is, if testimony purporting to be scientific is based on the expert’s *ipse dixit*—it cannot meet due process requirements of relevance and reliability. Nonetheless, the argument that future dangerousness testimony is inherently unreliable has been remarkably unsuccessful.⁷⁸ The overwhelming majority of courts that have addressed the issue since *Barefoot* have simply found its constitutionality beyond question. The Texas Court of Criminal Appeals in 1998, for example, addressing the scientific validity of future dangerousness testimony, found it to be reliable enough.⁷⁹

There have been however, a few muted stirrings of unease. For example, the Fifth Circuit acknowledged the issue of whether *Daubert* implicitly overruled *Barefoot*, although it declined to reach that issue, in *Tigner v. Cockrell*.⁸⁰ In addition, concurring in the Fifth Circuit Court of Appeals’ per curiam opinion in *Flores v. Johnson*,⁸¹ Judge Garza excoriated the Texas courts’ use of expert future dangerousness testimony (but found himself bound nonetheless by *Barefoot*).⁸² Further, as discussed above, in *United States v. Beckford*, the court declined to permit evidence of unadjudicated prior offenses to show future dangerousness on due process grounds.

In *Flores*, Judge Garza discussed in detail the psychological testimony used in Texas courts to support future dangerousness. The expert who testified in *Flores*, Dr. Griffith, was “frequently the state’s star witness” and had never once testified that any

⁷⁶ *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 591 (1993).

⁷⁷ *See Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146-47 (1997).

⁷⁸ *See WEBSTER ET AL.*, *supra* note 1, at 17 (observing that the political and legal pressures on experts to predict violence in the United States and Canada are overwhelming) (citing cases).

⁷⁹ *See Nenno v. State*, 970 S.W.2d 549 (Tex. 1998) (construing the Texas rules of evidence, which require a similar inquiry to that of *Daubert*), *overruled on other grounds by State v. Terrazas*, 4 S.W.3d 720 (Tex. Crim. App. 1988). “When addressing fields of study aside from the hard sciences, such as the social sciences or fields that are based primarily upon experience and training as opposed to the scientific method, [the] requirement of reliability applies but with less rigor than to the hard sciences.” *Nenno*, 970 S.W.2d. at 561.

⁸⁰ 264 F.3d 521 (5th Cir. 2001) (dismissing because a decision on collateral review would have violated the non-retroactivity principle).

⁸¹ 210 F.3d 456 (5th Cir. 2000) (per curiam).

⁸² *Id.* at 464 (Garza, J., concurring) (castigating the “inadequacy of the science” behind expert future dangerousness testimony); *see also supra* note 29.

defendant did not pose a future danger.⁸³ Judge Garza observed that it is still as true today as it was in *Barefoot's* time that “[n]either the Court nor the State of Texas has cited a single reputable scientific source contradicting the unanimous conclusion of professionals in this field that psychiatric predictions of long-term future violence are wrong more often than they are right.”⁸⁴ Moreover, when considered in juxtaposition with the strict admissibility requirements for most expert testimony—especially in civil trials, where the stakes are much lower—Dr. Griffith’s testimony became strikingly inadequate.⁸⁵

The *Barefoot* decision permitted experts to testify to future dangerousness because the state permitted such testimony by lay witnesses. Judge Garza pointed out that the problem with having an expert rather than a lay person testifying to future dangerousness is that it gives junk science the “imprimatur of scientific fact.”⁸⁶ Indeed, faced with such testimony, “juries are almost always persuaded.”⁸⁷ Although he acknowledged the jury’s right to impose death as an appropriate punishment for a vicious crime, Judge Garza concluded that “the legitimacy of our legal process is threatened” by allowing such testimony without any scientific validity into evidence.⁸⁸ Nonetheless, Flores’ death sentence was upheld.

Because juries give so much credence to expert testimony, and because the scientific literature evaluating the predictive value of clinical judgments about future violence has shown that such expert predictions are no better than lay judgments, the reliability standard articulated in *Daubert* and its progeny should apply to future dangerousness evidence. If *Daubert* standards were applied to the kinds of clinical predictions currently offered in our courts, they would not be admitted because they do not meet any of the criteria for scientific validity. This absence of scientific validity is of heightened concern in the context of capital sentencing hearings, where the jury hearing the evidence might very well

⁸³ *Flores*, 210 F.3d at 462 (Garza, J., concurring).

⁸⁴ *Id.* at 463 (Garza, J., concurring) (citing Justice Blackmun’s *Barefoot* dissent).

⁸⁵ *Id.* at 464 (Garza, J., concurring).

⁸⁶ *Id.* at 466 (Garza, J., concurring). Judge Garza opined that admitting a psychiatric prediction of dangerousness was akin to permitting a phrenologist—the example Justice Stevens used in *Joiner* of junk science—to testify that the bumps on a defendant’s skull could predict dangerousness. *Id.* at 465 n.12. He noted that “the phrenologists’ testimony appears no less scientific.”

⁸⁷ *Id.* at 466. The Court of Criminal Appeals acknowledged in the *Flores* case that, with only one exception, it was unable to find the evidence insufficient in any case where an expert had testified that the defendant posed a future danger. See *Flores v. Johnson*, 871 S.W.2d 714, 717 & n.3 (Tex. Crim. App. 1993) (en banc).

⁸⁸ *Flores*, 210 F.3d at 470 (Garza, J., concurring).

impose the death penalty. It is anomalous that more should be required to demonstrate relevance and reliability in civil than criminal cases.

III. EXPLAINING AND PREDICTING VIOLENCE

Despite an ocean of disparaging commentary on the use of future dangerousness testimony, it continues to find a place in court.⁸⁹ If, however, expert testimony were required to rest on empirically sound foundations, do such foundations exist? Given that current court predictions are woefully inaccurate, is it possible to improve them? This section discusses the latest research on the biology of violent behavior, which explains propensities toward violence in terms of brain structure and function. This section then outlines current research on violence prediction, contrasting clinical predictions—which, at best, are only modestly better than chance⁹⁰—with actuarial predictions, which appear to offer greater accuracy than any other method of prediction.

A. *The Biology of Violence*

Violent behavior is at least partly based on biology. Violent propensities have been shown to correlate with abnormalities in the structure and function of the brain and the central nervous system (CNS). These abnormalities, combined with environmental factors—such as stress or drug and alcohol abuse—can increase the chances that a particular individual will become violent. Indeed, many of the risk factors measured by actuarial assessments of violence risk (discussed later in this section) may be tied to an underlying biological function. Thus, biology may explain the statistical correlations between violence and risk factors.⁹¹

⁸⁹ See Mark D. Cunningham & Thomas J. Reidy, *Don't Confuse Me With the Facts: Common Errors in Violence Risk Assessment at Capital Sentencing*, 26 CRIM. JUST. & BEHAV. 20, 22 (1999).

⁹⁰ See Monahan, *supra* note 3, at 903 (noting two studies showing that the clinicians predicted violence with only a modest, though statistically significant, improvement over chance); see also Shuman & Sales, *supra* note 71, at 1228 (critiquing the use of clinical judgment testimony as lacking in scientific validity).

⁹¹ Correlation is the statistical degree of relationship between two variables. CHRIS SPATZ, BASIC STATISTICS: TALES OF DISTRIBUTIONS 77, 81 (5th ed. 1993) (defining correlation and explaining that the symbol for correlation—the correlation coefficient—is *r*). A variable is something that can be quantified and that exists in more than one

While we still do not know exactly how the brain works, we have a good idea about its capabilities, and there is a growing body of evidence on the biology of violence. Body, mind, and emotions are the products of evolutionary processes.⁹² The human brain processes information first through evolutionarily conserved systems, that is, old, instinctive kinds of systems that are first and foremost, “emotional.”⁹³ The newly evolved parts of the brain⁹⁴ get their information from these primordial, emotional systems.

Dysfunction in the lower brain centers, such as the amygdala,⁹⁵ can create problems at the emotion-forming stage of information processing.⁹⁶ On the other hand, dysfunction of the higher centers may create a lack of inhibition, allowing the release of primordial instinctive behaviors.⁹⁷ Any disruption in the

amount. *Id.* at 6. In simple terms, when there is a perfect correlation between two variables, $r = 1.00$; when there is no correlation, $r = 0$. *Id.* at 81. See *infra* Part IV.B.2 for a discussion of the risk factors measured by actuarial instruments.

⁹² See Jones, *supra* note 63, at 171 (noting that “because its neural architecture is genetically influenced, the brain’s processing pathways tend to reflect the cumulated result of hundreds of millions of years of natural selection”).

⁹³ See ANTONIO R. DAMASIO, *THE FEELING OF WHAT HAPPENS: BODY AND EMOTION IN THE MAKING OF CONSCIOUSNESS* 51 (1999). It has been suggested that the term “feeling” should be reserved for the mental experience of an emotion, while the term “emotion” should be used to designate the collection of responses, many of which are publicly observable. See *id.* at 42. It is in this context that we use the term emotion. Lower brain centers support our emotions through motor, autonomic, and endocrine responses that evolved as much for survival as for social signaling. In animals that have consciousness, emotion permeates the thought process and is devoted to the organism’s survival. The biological function of emotion is to (1) produce a specific reaction, and (2) regulate the internal state so it can be prepared for the specific reaction. See *id.* at 100.

It is these older, more primordial systems—which include such structures as the reticular activating system (RAS), hypothalamus, and limbic system—that control arousal levels and that first assess incoming sensory information in a “fight or flight” manner (kill or be killed, attack prey, flee a predator, immediate survival response). If no immediate threat exists, the information is assessed by these systems for benign adaptation (eat when hungry, drink when thirsty, instinctive survival response).

⁹⁴ These “higher” centers—which include the cerebral cortex, and are found in the frontal lobes of the brain—are our most highly evolved brain structures. They are in charge of critical judgment and learning. F.N. Dempster, *The Rise and Fall of the Inhibitory Mechanism: Toward a Unified Theory of Cognitive Development and Aging*, 12 *DEVELOPMENTAL REV.* 45 (1992).

⁹⁵ See *infra* text accompanying notes 109-13.

⁹⁶ See DAMASIO, *supra* note 93, at 13.

⁹⁷ If the cortex loses some of its inhibitory power, for example, by decreased blood flow to the frontal cortex—known as “hypofrontality”—the individual will be more prone to react to environmental stimuli with an exaggerated response. See, e.g., D.R. Weinberger, *On the Plausibility of “The Neurodevelopmental Hypothesis” of Schizophrenia*, 14 *NEUROPSYCHOPHARMACOLOGY* 1S-11S (1996) (finding that decreases in frontal lobe function in normal adolescents, coupled with increased levels of sex hormones, can lead to hyperresponsiveness to stimuli). Hypofrontality is present in a number of psychiatric and neurological disorders, such as schizophrenia, bipolar disorder, anxiety disorders, and major depression. See R.D. Skinner et al., *Reduced Sensory Gating of the P1 Potential in Rape Victims and Combat Veterans with Post*

information flow between higher and lower centers of the brain affects reasoning, decision-making, and behavior.⁹⁸

There is some evidence, which is not without its critics, suggesting that criminality and violence may have a genetic component.⁹⁹ Studies of petty criminals in Sweden and Denmark, for example, observed that the biological parents—but not the adoptive parents—had increased rates of criminality over the population base rate.¹⁰⁰ Genes are not deterministic of behavior, however. They may increase the likelihood (probability or risk) that a behavior will occur, but they do not cause it directly.¹⁰¹

Other studies suggest that structural dysfunction can contribute to violent behavior. Damage, decreased uptake of energy-producing glucose, reduced blood flow, and reduced function have all been observed in the frontal cortex of violent individuals and murderers. Studies using PET and SPECT, which are imaging devices designed to measure brain function by showing changes in the metabolism of energy-producing glucose, demonstrate that murderers have decreased glucose utilization in the frontal lobes compared to age- and gender-matched subjects.¹⁰² When these subjects were divided into two groups that either (1)

Traumatic Stress Disorder, 9 DEPRESSION & ANXIETY 122 (1999); see also Edgar Garcia-Rill, *Disorders of the Reticular Activating System*, 49 MED. HYPOTHESES 379 (1997) (reviewing study that showed hypofrontality in schizophrenic patients). Hypofrontality may also be responsible for the release of fixed action patterns, which are more complex than reflexes, but less hierarchical than such planned voluntary motor strategies as writing or speaking. See RODOLFO R. LLINAS, *I OF THE VORTEX: FROM NEURONS TO SELF* 134 (2001). These patterns can result in repetition of sequences of movements, e.g. stabbing repeatedly, and may be part of uncontrolled action patterns rather than rage. Hypofrontality has been shown to occur during dreaming, which may explain why we accept our dreams so readily, why we do not question that we are flying or embroiled in an unrealistic situation. See P. Maquet et al., *Functional Neuroanatomy of Human Rapid-Eye-Movement Sleep and Dreaming*, 383 NATURE 163 (1996).

⁹⁸ This connectivity between the higher and lower centers, so critical to brain function, will not be formed unless the central nervous system (CNS) is exposed to the environment. For this reason, the influence of the environment on violent behavior cannot be overlooked.

⁹⁹ See DEBRA NIEHOFF, *THE BIOLOGY OF VIOLENCE: HOW UNDERSTANDING THE BRAIN, BEHAVIOR, AND ENVIRONMENT CAN BREAK THE VICIOUS CIRCLE OF AGGRESSION* 238 (1999). For a critical perspective, see STEPHEN J. GOULD, *THE MISMEASURE OF MAN* 269 (2d ed. 1981).

¹⁰⁰ See NIEHOFF, *supra* note 99, at 238. Notably, these studies did not examine violent crime specifically, but included any criminal infractions as an outcome. See *id.* The theory underlying these studies was that if the environment was the important variable rather than inheritance, the rate should have been higher with both sets of parents.

¹⁰¹ See R. Plomin & M. Rutter, *Child Development, Molecular Genetics, and What to Do with Genes Once They Are Found*, 69 CHILD DEV. 1223, 1224 (1998).

¹⁰² See Adrian Raine et al., *Brain Abnormalities in Murderers Indicated by Positron Emission Tomography*, 42 BIOLOGICAL PSYCHIATRY 495 (1997); Adrian Raine et al., *Selective Reductions in Prefrontal Glucose Metabolism in Murders*, 36 BIOLOGICAL PSYCHIATRY 365 (1994) [hereinafter Raine, *Selective Reductions*].

had or (2) had not suffered early psychosocial deprivation—such as childhood abuse or neglect—it was the group without deprivation that had lower glucose metabolism in the frontal lobes.¹⁰³ This study was extended to a population of impulsive murderers¹⁰⁴ compared to a single predatory violent offender,¹⁰⁵ and the results suggest that impulsive murderers have decreased glucose metabolism in the frontal lobes while predatory murderers may not.¹⁰⁶ Additional studies have found abnormal temporal lobe metabolism in violent subjects.¹⁰⁷

Lesions of the frontal lobes have been associated with increases in the risk of aggressive and violent behavior.¹⁰⁸ There are reductions in the volumes of the frontal and temporal cortex in violent and antisocial personality disorder patients.¹⁰⁹ Reduced blood flow to the frontal lobes, known as “hypofrontality,” is evident in such disorders as schizophrenia, anxiety disorders, and depression.¹¹⁰ Interestingly, adolescents on average show a mild form of this condition.¹¹¹ The brain, especially the frontal lobes, is still maturing, so that the full-blown inhibition exercised by this

¹⁰³ See Adrian Raine et al., *Reduced Prefrontal and Increased Subcortical Brain Functioning Assessed Using Positron Emission Tomography in Predatory and Affective Murders*, 16 BEHAV. SCIENCES & LAW 319 (1998). However, in a study of a serial killer with at least forty-five victims, glucose utilization in the frontal lobes was perfectly normal. See NIEHOFF, *supra* note 99, at 111.

¹⁰⁴ Impulsive aggression is defined as “a response to physical or verbal aggression initiated by others that is relatively uncontrolled and emotionally charged.” Raine, *supra* note 103, at 320.

¹⁰⁵ Predatory aggression is defined by the author as “controlled, purposeful aggression lacking in emotion that is used to achieve a desired goal.” *Id.* One problem with this type of study is the subjective nature of its definitions.

¹⁰⁶ See Raine, *Selective Reductions*, *supra* note 102, at 370. A key problem here is that comparison of one group to another that has a single member cannot said to be anything more than suggestive of avenues for future research.

¹⁰⁷ See D. Seidenwurm et al., *Abnormal Temporal Lobe Metabolism in Violent Subjects: Correlation of Imaging and Neuropsychiatric Findings*, 18 AM. J. NEURORADIOLOGY 625, 631 (1997); H. Soderstrom et al., *Reduced Regional Cerebral Blood Flow in Non-Psychotic Violent Offenders*, 98 PSYCHIATRIC RES. IN NEUROIMAGING 29 (2000). The temporal lobe contains limbic structures like the hippocampus and amygdala, which are associated with emotional states.

¹⁰⁸ See J. Graffman et al., *Frontal Lobe Injuries, Violence, and Aggression: A Report of the Vietnam Head Injury Study*, 46 NEUROLOGY 1231 (1996).

¹⁰⁹ See H. Critchley et al., *Prefrontal and Medial Temporal Correlates of Repetitive Violence to Self and Others*, 47 BIOLOGICAL PSYCHIATRY 928 (2000); Adrian Raine et al., *Reduced Prefrontal Gray Matter Volume and Reduced Autonomic Activity in Antisocial Personality Disorder*, 57 ARCHIVES GEN. PSYCHIATRY 119 (2000).

¹¹⁰ See *supra* note 90. Overreactions to innocuous stimuli are likely in these disorders, particularly the anxiety disorders. Events or conditions could easily lead to exaggerated responses (panic attacks) or incapacitation (fear of leaving the house). Cornering such individuals or placing them under undue stress could elicit exaggerated reactions, striking out due to overperceived threats.

¹¹¹ See J.N. Giedd et al., *Brain Development During Childhood and Adolescence: A Longitudinal MRI Study*, 2 NATURE NEUROSCIENCE 861 (1999).

region in adulthood is not yet in play during adolescence.¹¹² MRI testing in adolescents shows mild decreases in frontal lobe function compared to adults, which, coupled with increased levels of sex hormones, can lead to hyperresponsiveness to stimuli.¹¹³ In general, these results, along with neurocognitive models of aggression and psychopathy,¹¹⁴ suggest that reduction of function in the frontal lobes may be responsible for impaired critical judgment and poor impulse control.

Studies on amygdalar function suggest that damage to this part of the brain might lead to disengagement and lack of empathy.¹¹⁵ Electrically stimulating the amygdala can elicit the physiological and behavioral signs of emotional states.¹¹⁶ On the other hand, if the amygdala is destroyed, the result is an unusually placid and emotionally unengaged individual.¹¹⁷ People and animals with this condition cannot muster the emotions which accompany fearful or painful situations. They are unable to recognize the context of danger although they are capable of emitting the right gestures.¹¹⁸ They may also be unable to empathize with others who do feel the emotions of fear and pain.¹¹⁹

Environmental factors often play a role in violent behavior. Animal studies have demonstrated that social isolation, for example, can produce aggression that intensifies as the isolation time increases, particularly when such isolation takes place at puberty.¹²⁰ Conversely, social *interaction* can lead to violent tendencies in persons who associate with peer groups that engage in deviant behavior.¹²¹ Alcohol consumption, which can effectively

¹¹² See *id.*; E.R. Sowell et al., *In Vivo Evidence for Post-Adolescent Brain Maturation in Frontal and Striatal Regions*, 2 NATURE NEUROSCIENCE 859 (1999).

¹¹³ See Weinberger, *supra* note 97, at 1S.

¹¹⁴ See R.J.R. Blair, *Neurocognitive Models of Aggression, Antisocial Personality Disorder, and Psychopathy*, 71 J. NEUROLOGY, NEUROSURGERY & PSYCHIATRY 727 (2001).

¹¹⁵ See NIEHOFF, *supra* note 99, at 92.

¹¹⁶ See *id.*

¹¹⁷ See ANTONIO R. DAMASIO, DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN 98-99 (1994). This is known as the Kluver-Bucy syndrome. See *id.*

¹¹⁸ See *id.*

¹¹⁹ See *id.*

¹²⁰ See D. Brunner & R. Hen, *Insights into the Neurobiology of Impulsive Behavior from Serotonin Receptor Knockout Mice*, 836 ANNALS N.Y. ACAD. SCI. 81 (1997); see also MERLIN DONALD, A MIND SO RARE: THE EVOLUTION OF HUMAN CONSCIOUSNESS 150 (2001) (finding that socially isolated humans do not develop language or any form of symbolic thought); L. Valzelli & S. Bernasconi, *Aggressiveness by Isolation and Brain Serotonin Turnover Changes in Different Strains of Mice*, 5 NEUROPSYCHOBIOLOGY 129 (1979) (finding that aggression in isolated rats did not become manifest until after puberty).

¹²¹ See Thomas J. Dishion et al., *Peer Group Dynamics Associated with Iatrogenic Effects in Group Interventions with High Risk Young Adolescents*, 91 NEW DIRECTIONS IN

diminish the activity of small brain cells responsible for cortical function, figures into two out of every three violent crimes.¹²²

Physiological responses to environmental stimuli can also predispose individuals to engage in violent behavior. For example, some violent criminals have lowered CNS¹²³ and autonomic nervous system (ANS) arousal.¹²⁴ Measures of antisocial behavior in fifteen year-old males have also been correlated with reduced ANS activation.¹²⁵ Further studies have shown that measures of underarousal of the CNS and ANS taken at fifteen years of age were related to criminality status assessed at twenty-four years of age.¹²⁶

Hormones, which are regulated by the hypothalamus and the ANS, can play a role as well. Testosterone in boys surges at age ten, rising to a plateau by age fourteen, when aggressive behavior starts accelerating.¹²⁷ However, delinquent behavior, adjustment problems, and rebelliousness are actually more likely to be associated with *lower* testosterone levels.¹²⁸ Studies of recidivistic violent offenders, adults with antisocial personality disorder, and antisocial adolescents have all documented statistically significant reductions in levels of cortisol, the main circulating stress hormone.¹²⁹ This suggests that psychopathy¹³⁰ may involve a stress

CHILD & ADOLESCENT DEV. 79 (2001) (reviewing study finding that boys who did not drink or smoke before the age of thirteen or fourteen, but who befriended boys who did, advanced in statistically predictable ways to abuse drugs two years later); *see also* E. Marshall, *The Shots Heard Around the World*, 289 SCI. 570, 573 (2000) (suggesting that aggregating delinquents in group homes or sending them to prison is graduate school for violence).

¹²² See J. Roizen, *Issues of Epidemiologist of Alcohol and Violence*, in ALCOHOL AND INTERPERSONAL VIOLENCE: FOSTERING MULTIDISCIPLINARY PERSPECTIVES 3-36 (Susan E. Martin ed., 1993).

¹²³ The normal human CNS displays immediate, short-term, instinctive reflexive activity as a first line of defense to real or perceived threats.

¹²⁴ See NIEHOFF, *supra* note 99, at 181. Lowered levels of arousal were accompanied by decreased activation of the reticular activating system (RAS), which is the part of the brain that controls sleep/wake cycles and arousal, and lowered hypothalamic-modulated stress responses. *See id.* Generally speaking, the hypothalamus—along with the RAS—helps regulate the body's physiological response to stress, often referred to as "fight or flight." Robert M. Sapolsky, *The Stress Response and the Emergence of Stress-Related Disease*, in STRESS, THE AGING BRAIN, AND THE MECHANISMS OF NEURON DEATH 3-9 (Robert M. Sapolsky ed., 1999).

¹²⁵ See Adrian Raine et al., *Autonomic Nervous System Factors Underlying Disinhibited, Antisocial, and Violent Behavior*, 794 ANNALS N.Y. ACAD. SCI. 46 (1996) (reviewing study).

¹²⁶ See Adrian Raine et al., *Relationship Between Central and Autonomic Measures of Arousal at Age 15 Years and Criminality at Age 24 Years*, 47 ARCHIVES GEN. PSYCHIATRY 1003 (1990).

¹²⁷ See NIEHOFF, *supra* note 99, at 159.

¹²⁸ *See id.* at 160.

¹²⁹ See B. Bergman & B. Brismar, *Hormone Levels and Personality Traits in Abusive and Suicidal Male Alcoholics*, 18 ALCOHOLISM: CLINICAL & EXPERIMENTAL RES. 311

response disorder, one in which the response to a threat is too weak rather than too strong. Psychopaths appear to lack a conscience, but what they may lack is the biological machinery necessary to warn them that they are heading for disaster.¹³¹

Current research on the biology of violence suggests that violence is pathological, that it is “normal” aggression gone awry.¹³² Violent tendencies can develop in early infancy,¹³³ or can emerge after the onset of puberty.¹³⁴ Adolescence (that is, up to the age of twenty-five) is not only the stage at which most violent crimes are committed, it is also the stage at which hormone levels are fluctuating widely.¹³⁵ The brain, especially the frontal lobes, is still maturing during adolescence.¹³⁶ In addition, normal adolescents appear to have a sensory gating deficit—an overresponsiveness to repetitive or inconsequential stimuli—which also leads to increased distractability.¹³⁷ Further, the brain structure changes during adolescence: excess synapses in the brain that may have

(1994); B. McBurnett et al., *Anxiety, Inhibition, and Conduct Disorder in Children*, 30 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 192 (1991); M. Virkunen, *CSF Biochemistries, Glucose Metabolism, and Diurnal Activity Rhythms in Alcoholic, Violent Offenders, Fire Setters, and Healthy Volunteers*, 51 ARCHIVES GEN. PSYCHIATRY 20 (1994); M. Virkunen, *Urinary Free Cortisol in Habitually Violent Offenders*, 72 ACTA PSYCHIATRICA SCANDINAVICA 40 (1985).

¹³⁰ It is hard to define psychopathy. The problem is that no one is sure exactly what a psychopath is. For example, Harris et al. describe psychopathy as “a lifelong persistent condition characterized . . . by aggression beginning in early childhood, impulsivity, resistance to punishment, general lack of emotional attachment or concern for others, dishonesty and selfishness in social interactions, and high levels of promiscuous and uncommitted sexual behavior.” Grant T. Harris et al., *The Construct of Psychopathy*, 28 CRIME & JUST. 197, 197 (2001). The research of Raine et al., on the other hand, identifies two kinds of psychopathy, impulsive and controlled. See Raine, *supra* note 125, at 46.

¹³¹ See NIEHOFF, *supra* note 99, at 181.

¹³² See R.J. Davidson et al., *Dysfunction in the Neural Circuitry of Emotion Regulation a Possible Prelude to Violence*, 289 SCI. 591, 591 (2000). Aggression is thought to be part of the normal repertoire of behaviors that has arisen to balance the need of the individual to look out for himself and still maintain good standing within the group. See NIEHOFF, *supra* note 99, at 76.

¹³³ See F.B.M. De Waal, *Primates: A Natural Heritage of Conflict Resolution*, 289 SCI. 586, 586 (2000).

¹³⁴ See, e.g., NIEHOFF, *supra* note 99, at 159 (1999).

¹³⁵ See *supra* text accompanying note 121.

¹³⁶ See *supra* notes 106-07 and accompanying text.

¹³⁷ See Lisa Rasco et al., *Effects of Age on Sensory Gating of the Sleep State-Dependent P1/P50 Midlatency Auditory Evoked Potential*, 3 SLEEP RES. ONLINE 97 (2000), at <http://www.sro.org/bin/article.dl?Paper&1930&0&0>. Sensory gating has been conceptualized as a critical function of the CNS to filter out extraneous information and to focus on newer, more salient stimuli. Adding this distractability to a stressful environment or other condition which increases arousal, e.g. hormones, probably increases the chances of exaggerated responsiveness, too much “fight,” or too much “flight.” For a detailed discussion of the stress response, see Edgar Garcia-Rill & Erica Beecher-Monas, *Gatekeeping Stress: The Science and Admissibility of Post-Traumatic Stress Disorder*, 24 U. ARK. LITTLE ROCK L. REV. 9, 12-14 (2001).

developed in pre-teen growth spurts begin to be eliminated. All of these brain changes, in conjunction with widely fluctuating hormone levels, may explain why adolescents are at much greater risk for engaging in violent behavior than any other age group.¹³⁸

Hypofrontality, associated with Axis I psychiatric disorders,¹³⁹ may explain why some individuals who are experiencing active symptoms of mental illness are at greater risk for committing acts of violence.¹⁴⁰ Hypofrontality, when caused by drug or alcohol use, may further explain why substance abuse is also a risk factor for violence.¹⁴¹ Not surprisingly, some killers have demonstrably decreased blood flow to the frontal lobes, and most violent offenders lack the cognitive and control capacities of normal adults.¹⁴²

Studies indicating that there may be two distinct groups of violent offenders¹⁴³ suggest that not all violence originates from the same sources within the brain. These groups are distinguished by the type of aggression they display. One group exhibits “impulsive aggression,” which is “a response to physical or verbal aggression initiated by others with violence that is relatively uncontrolled and emotionally charged.”¹⁴⁴ Others exhibit “predatory aggression,” which is “controlled, purposeful aggression lacking in emotion that is used to achieve a desired goal.”¹⁴⁵ Some interesting research shows that offenders who display predatory aggression may suffer from an under-active stress response,¹⁴⁶ while those exhibiting impulsive violence suffered from an over-active stress response.¹⁴⁷

¹³⁸ “Young age at time of first violent incident” is a risk factor measured by the HCR-20. See CHRISTOPHER D. WEBSTER ET AL., *THE HCR-20 SCHEME: THE ASSESSMENT OF DANGEROUSNESS AND RISK* (1995); see also U.S. DEP’T OF JUSTICE, *VIOLENT CRIME 1* (Apr. 1994) (reporting that people between the ages of sixteen and twenty-four consistently commit the highest percentage of violent crimes), available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/viocrm.pdf>.

¹³⁹ See generally AMERICAN PSYCHIATRIC ASSOCIATION, *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS* (4th ed. 1994).

¹⁴⁰ The HCR-20 lists active symptoms of major mental illness as a risk factor. See *infra* notes 214-20 and accompanying text. Likewise, the VRAG lists meeting the DSM criteria for schizophrenia as a risk factor. See *infra* text accompanying note 229. There is however, research showing that the mentally ill as a group are no more violent than the general population.

¹⁴¹ See *supra* text accompanying note 116; *infra* note 200 and text accompanying note 211.

¹⁴² See Paul H. Robinson, *Preventing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114 HARV. L. REV. 1429 (2001).

¹⁴³ See Raine, *supra* note 103, at 320.

¹⁴⁴ *Id.*

¹⁴⁵ *Id.*

¹⁴⁶ See *supra* notes 123-25 and accompanying text.

¹⁴⁷ See Raine, *Selective Reductions*, *supra* note 102, at 2 (finding that impulsive murderers have decreased glucose metabolism in the frontal lobes while predatory murderers do not).

This research—particularly the studies on predatory violent offenders—may suggest why psychopathy is a risk factor for future violence.¹⁴⁸

Other risk factors—those one could call sociological rather than biological—such as family situation, poverty, high-crime neighborhoods, and urban residence may also affect stress levels, thereby leading to biological repercussions. Current research in neuroscience indicates that violent behavior is a complex intermingling of nature and nurture. It may be that “neuropsychological impairments disrupt normal development and increase vulnerability” to poor social environments.¹⁴⁹

The brain is a complex system, understandable only in a statistical sense. Complex systems can only be considered in a statistical manner because they are probabilistic, not certain.¹⁵⁰ With this in mind, risk assessment could perhaps provide the critical link between biology and the probability that violence will eventually manifest itself in a given individual.

B. *Assessing Risk: A New Paradigm for Predicting Violence*

In a parallel development with the growth of knowledge about brain structure and function (and largely isolated from it) is the burgeoning field of risk assessment.¹⁵¹ The idea of risk encompasses notions of statistical probability, factor analysis, and the likelihood of event occurrence.¹⁵² The idea of risk with respect to violent behavior has a number of advantages over the notion of future dangerousness. Risk shifts the focus from a yes/no analysis to a probabilistic statement and requires balancing the seriousness of the possible outcome with the probabilities of their occurrence based on specific risk factors.¹⁵³ The key variables in risk assessment of any stripe are the outcomes, their probabilities of

¹⁴⁸ See *infra* notes 194, 199 and text accompanying note 211.

¹⁴⁹ James C. Howell & J. David Hawkins, *Prevention of Youth Violence*, 24 CRIME & JUST. 263, 268 (1998) (citing Terrie E. Moffitt, *Adolescence Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy*, 100 PSYCH. REV. 674 (1993)).

¹⁵⁰ See STUART KAUFFMAN, *AT HOME IN THE UNIVERSE: THE SEARCH FOR THE LAWS OF SELF-ORGANIZATION AND COMPLEXITY* 18 (1995).

¹⁵¹ See Monahan, *supra* note 3, at 914. Something is a risk factor if it precedes the outcome (violence, in this case) and correlates with the outcome. Helena Kraemer et al., *Coming to Terms with the Terms of Risk*, 54 ARCHIVES GEN. PSYCHIATRY 337 (1997).

¹⁵² See Jonathan Simon, *Managing the Monstrous: Sex Offenders and the New Penology*, 4 PSYCHOL. PUB. POL'Y & L. 452, 452 (1998).

¹⁵³ See David Carson, *A Risk-Management Approach to Legal Decision-Making About 'Dangerous' People*, in *LAW AND UNCERTAINTY: RISKS AND LEGAL PROCESS* 20 (Robert Baldwin ed., 1997) (discussing the paradigm shift from dangerousness to risk).

occurrence, the uncertainty about the outcomes, and time.¹⁵⁴ Risk assessment is supposed to objectively quantify the probabilities and consequences of adverse events by looking at physical and natural processes.¹⁵⁵ Risk is a central concern of insurance, business, government regulation (including environmental legislation), and health. Risk management is increasingly seen as the provenance of penology.¹⁵⁶ Criminal justice is becoming actuarial, and the use of statistics to conceive group boundaries is gaining prominence.¹⁵⁷

When we speak of predicting violent behavior, we are assessing the risk that an individual will become violent at some later date. Although various risk factors have been analyzed and numerous studies corroborate their relationship to violent behavior, that is merely the beginning of risk analysis. No one of these factors, standing alone, has significant predictive power; they must be analyzed in concert.¹⁵⁸ One of the significant problems with clinical judgment is an apparent inability to adjust predictions according to these interrelationships.¹⁵⁹

1. Clinical Predictions

A number of studies about human decisionmaking illustrate why clinical judgments about future dangerousness tend to be error prone and unreliable.¹⁶⁰ Experts, like lay people, tend to use a number of flawed heuristics to simplify decisionmaking.¹⁶¹ These

¹⁵⁴ See K.R. MCCRIMMON & D.A. WEHRUNG, *TAKING RISKS: THE MANAGEMENT OF UNCERTAINTY* (1988).

¹⁵⁵ See Paul Slovic, *Trust, Emotion, Sex, Politics, and Science: Surveying the Risk Assessment Battlefield*, 1997 U. CHI. LEGAL F. 59, 63 (discussing risk analysis and differences among population groups in risk assessment).

¹⁵⁶ See Simon, *supra* note 152, at 452 (discussing the transformation of penology from seeing crime as a problem of molding offenders to community norms to "managing high-risk categories and subpopulations").

¹⁵⁷ For example, a number of risk assessment tools are used in making decisions about parole. See M. Dolan & M. Doyle, *Violence Risk Prediction: Clinical and Actuarial Measures and the Role of the Psychopathy Checklist*, 177 BRIT. J. PSYCH. 303, 304 (2000) (listing assessment instruments used for parole determinations).

¹⁵⁸ See John Monahan, *Clinical and Actuarial Predictions of Violence*, in 1 DAVID FAIGMAN ET AL., *supra* note 12, § 7-2.1.1 (Supp. 2000) (noting that "it is crucial for future studies to use multiple measures of violence rather than the single measures that have characterized most prior research").

¹⁵⁹ See QUINSEY, *supra* note 2, at 56.

¹⁶⁰ See Shuman & Sales, *supra* note 71, at 1228-29 (discussing heuristics and biases in clinical judgment).

¹⁶¹ See DANIEL KAHNEMAN ET AL., *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* (1982). A base rate is the rate at which a particular outcome (such as violence) occurs throughout the population. See Cunningham & Reidy, *supra* note 89, at

flawed heuristics contribute to the high error rate.

One of these flawed heuristics is the representativeness heuristic. Clinical decisionmakers tend to erroneously assume the representativeness of events by ignoring sample sizes and base rates.¹⁶² In addition, clinicians, like other people, tend to think they have more information than they really do.¹⁶³ Clinicians (again, like other people) are poor at making extreme judgments; they tend to have more confidence in predictive variables with extreme values than is warranted.¹⁶⁴ In other words, although it is more difficult to predict statistically rare events (like violence), clinical judgments frequently fail to take that into account.¹⁶⁵

Ignoring base rates is a particular problem in predicting violence where the base rate of violent behavior is low overall and varies among different population sub-groups.¹⁶⁶ These cognitive phenomena are particularly troubling when the task is predicting violent behavior, and clinicians tend to use these error-inducing heuristics as often as lay people in making these assessments.¹⁶⁷ Clinicians, like other people, tend to be swayed by their stereotypes and prejudices, against which they measure the patient's behavior and which, because they have no relevance to medical expertise, may taint the decision.¹⁶⁸

A further problem with clinical judgments is that there is no way to assess the individual error rate of the particular expert proffering a future dangerousness opinion.¹⁶⁹ There is no way to know how many times the expert has opined that someone was dangerous when he was not (or vice-versa). Clinical judgment is

23.

¹⁶² See Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 COGNITIVE PSYCHOL. 430 (1972). These shortcuts are not consciously employed, but operate on a subliminal level to affect decisionmaking. *Id.*

¹⁶³ See QUINSEY, *supra* note 2, at 56.

¹⁶⁴ See H. Einhorn & R. Hogarth, *Confidence in Judgment: Persistence of the Illusion of Validity*, 85 PSYCHOL. REV. 395 (1978).

¹⁶⁵ See JOHN W. PARRY ET AL., AM. BAR ASS'N COMM'N ON MENTAL HEALTH AND PHYSICAL DISABILITY LAW, NATIONAL BENCHBOOK ON PSYCHIATRIC AND PSYCHOLOGICAL EVIDENCE AND TESTIMONY 20 (1998) (noting that "it is difficult to predict with certainty occurrences of statistically rare events").

¹⁶⁶ See *id.* at 223 (explaining how the statistical base-rate problem affects predictions of dangerousness); QUINSEY, *supra* note 2, at 60.

¹⁶⁷ See QUINSEY, *supra* note 2, at 62 (in a study assessing predictions of violence, "lay persons and the clinicians had few differences of opinion"); Shuman & Sales, *supra* note 71, at 1228 (noting that "expert judgments that are clinically derived, as opposed to actuarially derived, are as susceptible to error as lay judgements").

¹⁶⁸ See WEBSTER ET AL., *supra* note 1, at 28 n.5 (quoting Judge Bazelon).

¹⁶⁹ See *id.* at 25 (noting the problem of illusory correlations when assessors—who usually have no information about the accuracy of their predictions—learn of a violent action by a previously assessed patient, making that instance stand out and giving the clinicians a mistakenly optimistic view of their own prowess).

thus virtually untestable.¹⁷⁰ Reliance on personal experience can be much improved if it is augmented by statistically analyzed data, and that is the insight behind the development of a number of actuarial instruments for predicting human violence.

Repeated studies of actuarial methods have demonstrated them to be superior to clinical judgment standing alone.¹⁷¹ Even such instruments however, with their structured reasoning requirements, do not obviate all the problems of human judgment. For example, the risk factor descriptions may be vague, decreasing their reliability.¹⁷² Sometimes the factors are not independent, for example anger and the inability to sustain relationships.¹⁷³ Moreover, the time period that the risk assessment is to cover, the circumstances under which it will be implemented (prison for life, in the case of capital sentencing, and the perhaps eventual release into the community after serving a minimum term of the life sentence),¹⁷⁴ and the individual's motivation to refrain from violence (including motivation to comply with treatment), must all be taken into consideration in assessing risk.¹⁷⁵ Yet rarely are they addressed, either in court or in the actuarial instruments. A major problem with each of the risk instruments is its failure to correlate the risk factors with neuroscience and evolutionary biology.¹⁷⁶

Finally, it is important to bear in mind that risk is a social construct. Although it uses probabilistic analysis and quantification, it is not an exact science.¹⁷⁷ Indeed, all science is value-laden, and risk assessment is not different in that regard.¹⁷⁸ Using structured analysis however, offers many advantages in human decisionmaking, particularly in light of the difficulty people have in synthesizing differently weighted likelihoods of varying significance (such as risk factors for violent behavior).¹⁷⁹ Thus,

¹⁷⁰ See Shuman & Sales, *supra* note 71, at 1227 (noting the problem of unvalidated theories and skills).

¹⁷¹ See Cunningham & Reidy, *supra* note 89, at 28 (citing studies).

¹⁷² See Carson, *supra* note 153, at 258 (noting the problem of reliability). For example, even trained clinicians may differ on what exactly is meant by "glibness" (a factor on the PCL-R) or "lack of insight" (a factor on the VRAG and HCR-20).

¹⁷³ See *id.*

¹⁷⁴ If permitted by statute; in some states, life term is without parole.

¹⁷⁵ See *id.* at 261.

¹⁷⁶ But see Harris et al., *supra* note 130, at 197 (attempting to correlate the PCL-R with neuroscience and conceding that much research still needs to be done).

¹⁷⁷ See ROYAL SOCIETY, *RISK: ANALYSIS, PERCEPTION AND MANAGEMENT* 7 (1992) (explaining that some subjectivity is always a part of risk assessment).

¹⁷⁸ See Erica Beecher-Monas, *The Heuristics of Intellectual Due Process: A Primer for Triers of Science*, 75 N.Y.U. L. REV. 1563 (2000).

¹⁷⁹ See J. RICHARD EISER & JOOP VAN DER PLIGT, *ATTITUDES AND DECISIONS* 100 (1988) (observing that human decision "accuracy declines considerably when the number of features or the number of alternatives increases . . . [and] reliability with which choice

actuarial instruments may offer a distinct advantage in assessing risk of violent behavior.¹⁸⁰

2. Actuarial Instruments

One way to counter human cognitive error is to create a scoring instrument that takes into account the interrelationship of various risk factors, the population base rates, and assigns weights to the individual risk factors. Human judgment is not abandoned, it is simply structured into a formal reasoning process.¹⁸¹ Although there is some evidence that a multi-disciplinary team may be able to rival the accuracy of actuarial instruments, empirical data demonstrates that such structured analysis improves decisionmaking considerably.¹⁸² Moreover, in a legal system that places some value on cost/benefit analyses, actuarial instruments are not only more accurate, they are also more cost-effective.

Actuarial instruments are risk assessment tools that combine a number of risk factors in order to achieve an overall "score" that ranks levels of risk.¹⁸³ Currently, there are three prominent instruments used in assessing violence, the Violence Risk Assessment Guide (VRAG),¹⁸⁴ the Historical/ Clinical Risk Management 20-item scale (HCR-20),¹⁸⁵ and the Psychopathy Checklist (Revised) (PCL-R).¹⁸⁶ A number of studies of these instruments have shown that actuarial measures are more accurate predictors than clinical judgment.¹⁸⁷ Occasionally, these actuarial studies are finding their way into court.¹⁸⁸ The question is whether

rules are used tends to decrease as the decision-maker's information load increases").

¹⁸⁰ See N. Morris & M. Miller, *Predictions of Dangerousness*, in CRIME AND JUSTICE: AN ANNUAL REVIEW OF RESEARCH 1-50 (M. Tonry & M. Miller eds., 1985).

¹⁸¹ See QUINSEY, *supra* note 2, at 65; Dolan & Doyle, *supra* note 157, at 304 (observing that "[s]tructured clinical judgment represents a composite of empirical knowledge and clinical/professional expertise").

¹⁸² See J. Fuller & J. Cowan, *Risk Assessment in a Multidisciplinary Forensic Setting: Clinical Judgement Revisited*, 10 J. FORENSIC PSYCH. 276 (1999).

¹⁸³ See Monahan, *supra* note 3, at 903 (evaluating risk assessment instruments).

¹⁸⁴ See QUINSEY, *supra* note 2, at 141.

¹⁸⁵ See CHRISTOPHER D. WEBSTER ET AL., HCR-20: ASSESSING RISK OF VIOLENCE VERSION 2 (1997).

¹⁸⁶ See R. D. HARE, MANUAL FOR THE HARE PSYCHOPATHY CHECKLIST-REVISED (1991).

¹⁸⁷ See Dolan & Doyle, *supra* note 157, at 303 (citing studies).

¹⁸⁸ In *United States v. Barnette*, the prosecution expert used the PCL-R to substantiate his opinion that the defendant posed a future danger. 211 F.3d 803 (4th Cir. 2000). The court upheld the admissibility of this testimony under *Daubert*, although it declined to decide whether a *Daubert* analysis was required. *Id.* at 815. The court found that because the prosecution expert had based his opinion on the PCL-R together with the fourth edition *Diagnostic and Statistical Manual* (DSM-IV), produced by the American

they offer a more accurate means of assessing future dangerousness.

a. PCL-R

The PCL-R is an instrument that is designed to measure a “psychological construct” known as “psychopathy.”¹⁸⁹ No one is quite sure what psychopathy means,¹⁹⁰ but researchers seem to agree that the PCL-R can be used to predict recurring violence.¹⁹¹ The goal of the instrument is to predict violence in offenders and psychiatric patients by assessing “a constellation of affective, interpersonal and behavioural characteristics.”¹⁹² The instrument consists of twenty risk factors, scored on a three-point scale, with a score of greater than 30 (on a 40-point scale) indicating a strong propensity for violence.¹⁹³ The twenty factors are: Glibness/superficial charm; grandiose sense of self-worth; need for stimulation/ proneness to boredom; pathological lying; conning/manipulative; lack of remorse or guilt; shallow affect; callous/ lack of sympathy; parasitic lifestyle; poor behavioral controls; promiscuous sexual behavior; early behavioral problems; lack of realistic long-term goals; impulsivity; irresponsibility; failure to accept responsibility; many short-term marital relationships; juvenile delinquency; revocation of conditional release; criminal versatility.¹⁹⁴

Psychiatric Association, and with “observations of Barnette’s behavior; the actuarial approach; and the research on predicting future dangerousness” that it met *Daubert’s* reliability standard. *Id.* at 816. The court was careful to point out that just because testimony is admissible does not mean it cannot be attacked under cross-examination. *Id.*

¹⁸⁹ Harris et al., *supra* note 130, at 197 (2001).

¹⁹⁰ See *id.* at 197-98 (asserting that “[p]sychopathy is a real phenomenon . . . characterized, in males at least, by aggression beginning in early childhood, impulsivity, resistance to punishment, general lack of emotional attachment or concern for others, dishonesty and selfishness in social interactions, and high levels of promiscuous and uncommitted sexual behavior . . . mediated . . . by genes that modulate some neuroanatomical structures an monoamine oxidase-type A (MAO) neurotransmitters”) (no citations given).

¹⁹¹ See Darryl G. Kroner & Jeremy F. Mills, *The Accuracy of Five Risk Appraisal Instruments in Predicting Institutional Misconduct and New Convictions*, 28 CRIM. JUST. & BEHAV. 471, 473 (2001) (asserting that “the PCL-R has been shown to have utility in predicting criminal behavior” and citing studies).

¹⁹² Robert D. Hare, *The Hare PCL-R: Some Issues Concerning its Use and Misuse*, 3 LEGAL & CRIM. PSYCH. 99, 100 (1998).

¹⁹³ It has been supplemented for institutional screening purposes by the PCL-SV, which consists of only twelve items, with scores ranging from 0-24, and a cut-off score for violence above 18. The twelve items for the PCL-SV are: Superficial; grandiose; manipulative; lacks remorse; lacks empathy; does not accept responsibility; impulsive; poor behavioral controls; lacks goals; irresponsible; adolescent antisocial behavior; adult antisocial behavior. See *id.*

¹⁹⁴ *Id.* at 306, tbl. 1.

These factors are of two types: one that describes interpersonal dimensions (such as shallowness, callousness, dishonesty), and one that describes behavior (such as juvenile delinquency, criminal versatility, impulsivity), and the factors overlap somewhat with the American Psychiatric Association's concept of antisocial personality disorder.¹⁹⁵ Although these factors appear to be rather subjective, Hare argues that "the scoring criteria for each item . . . are explicit, and the meaning of an item (e.g., shallow affect) is based on these criteria, not on what the title of the item might mean to an individual clinician or researcher."¹⁹⁶

The information in the scoring instrument is derived from the subject's entire lifetime, so it cannot measure changes in symptoms (in response to treatment, for example).¹⁹⁷ The base rate in the offender population (those who have committed violent acts and those who have not) for violent propensities defined according to the PCL-R is fifteen to twenty-five percent in North America.¹⁹⁸ That is, fifteen to twenty-five percent of the general prison population is characterized as violent according to this instrument.¹⁹⁹ Both the VRAG and the HCR-20 incorporate the PCL-R, although the VRAG eliminated the clinical assessment component (and thus the subjectivity) and garnered the information strictly from the subject's file.²⁰⁰

One of the major problems with this instrument is that it is simply not available to anyone who wishes to examine it.²⁰¹ Thus, it fails one of the important criteria of falsifiability, the willingness to submit to critique (including interdisciplinary critique).²⁰² It is designed to be administered by specifically trained individuals on the basis of a semi-structured interview and file information (although it can be based on file information alone).²⁰³ In addition,

¹⁹⁵ See Harris et al., *supra* note 130, at 204 (noting that Hare has since expanded the concept of psychopathy from two to three factors, but that this expansion is largely untested).

¹⁹⁶ Hare, *supra* note 192, at 109.

¹⁹⁷ See *id.* at 116.

¹⁹⁸ See *id.* at 101 (noting the consistency of the base rate across differing populations and asserting that there is no evidence the test is biased due to race or mental disorder).

¹⁹⁹ See *id.* at 116.

²⁰⁰ See Kroner & Mills, *supra* note 191, at 480.

²⁰¹ See Hare, *supra* note 192, at 108-13 (arguing that only professionals trained in "the Hare PCL-R training program" should be allowed access to the instrument, particularly in litigation, and castigating judges who have made the instrument part of the public record—until threatened with copyright suits—and lawyers who have disseminated the instrument to clients).

²⁰² Beecher-Monas, *supra* note 178.

²⁰³ See Hare, *supra* note 192, at 101, 110 (stating that individuals must be "qualified to purchase or use" the Hare instrument). Notably, because only people who have gone

a number of researchers have criticized the PCL-R because some of the scoring variables on the checklist are not independent—that is, they are confounded.²⁰⁴ For example, some of the checklist items—like pattern of past violence—are linked to the outcome variable (violence), and this necessitates either statistical control for past criminal activity or removing confounding items from the checklist.²⁰⁵ However, when testing whether the instrument simply measured “a lengthy history of officially recorded criminal conduct,”²⁰⁶ researchers concluded that it had a “unique effect.”²⁰⁷ In other words, it does something more than just measure past patterns of violent conduct.

Nevertheless, even the instrument’s author recognizes its subjective nature and recommends that “at least two independent ratings be obtained and averaged.”²⁰⁸ People with a vested interest in the outcome are told that they should not administer the test (which makes one wonder about its suitability for litigation purposes).²⁰⁹ Moreover, attempts to define terms like “psychopath” have been weak.²¹⁰ Further, there is some question about its applicability to United States populations, because efforts at evaluating the PCL-R have been conducted almost entirely in Canada, where the minority population is much smaller. At least one recent study has challenged the PCL-R’s applicability to African-Americans, concluding that there are substantial differences in population groups with respect to the distribution of psychopathy scores, the relation of psychopath to impulsiveness, and the relevance of the underlying factors.²¹¹ Despite these major problems, many researchers consider the instrument to have at

through Hare’s qualification program may have access to the instrument, this isolates the Hare instrument from the robust critique that Karl Popper advocated as a prerequisite to valid science.

²⁰⁴ Confounders are unobserved variables that are correlated with the independent and dependent variables. See David H. Kaye, *The Dynamics of Daubert: Methodology, Conclusions and Fit in Statistical and Econometric Studies*, 87 VA. L. REV. 1933, 1998 (2001).

²⁰⁵ See Dolan & Doyle, *supra* note 157, at 305.

²⁰⁶ Harris et al., *supra* note 130, at 199.

²⁰⁷ When the researchers entered the “four best predictors reflecting criminal history” first into a multiple regression analysis, they found that the instrument still “showed a unique effect of psychopathy in the prediction of violent recidivism.” *Id.* at 199. The researchers do not quantify their findings however, so it is impossible to tell how unique the effect was, or what conduct was being measured.

²⁰⁸ Hare, *supra* note 192, at 113.

²⁰⁹ See *id.* at 116.

²¹⁰ See WEBSTER ET AL., *supra* note 1, at 27 n.3.

²¹¹ See Robert T. Salekin et al., *A Review and Meta-Analysis of the Psychopathy Checklist and the Psychopathy Checklist-Revised: Predictive Validity of Dangerousness*, 3 CLIN. PSYCH. 203, 208 (1996).

least some predictive value for community violence,²¹² and it is frequently used in prisons to determine eligibility for parole. It was also used by the prosecution expert in *Barnett* in the capital sentencing phase, and may see increased use in the courts. In comparison studies, however, the VRAG (which uses the PCL-R as one of its risk factors but without the clinical interview) had considerably better prediction rates.²¹³

b. HCR-20

This instrument is a scoring device developed to predict violent behavior in criminal and psychiatric populations.²¹⁴ It consists of twenty factors, ten taken from the subject's history (two of which address personality dysfunction); five clinical factors; and five risk management factors.²¹⁵ The information is gathered from clinical interviews, reports from family members, and file records.²¹⁶ In two studies the historical factors showed better correlation with violent outcomes than the clinical factors.²¹⁷ In a retrospective study of violent offenders, the ten historical factors performed better than the VRAG.²¹⁸ This makes sense because in a study of predictive factors for violent behavior, using a meta-analysis of sixty-four separate studies employing twenty-seven predictors, Bonta et al. found that the largest effects came from criminal history variables, followed by personal demographics, deviant lifestyle, and lastly, clinical variables.²¹⁹ The HCR-20 and the PCL-R have been shown to perform similarly in predicting post-release behavior in a number of tests.²²⁰

²¹² See John Monahan et al., *Developing a Clinically Useful Actuarial Tool for Assessing Violence Risk*, 176 BRIT. J. PSYCH. 312, 320 (2000). "Some" predictive value is not quantified by the author. See *id.*

²¹³ See Dolan & Doyle, *supra* note 157, at 307; see also Kroner & Mills, *supra* note 191, at 481.

²¹⁴ See WEBSTER, *supra* note 138.

²¹⁵ See *id.* at 27. The historical factors are previous violence; young age at first violent incident; relationship instability; employment problems; substance use problems; major mental illness; psychopathy; early maladjustment; personality disorder; prior supervision failure. See *id.* at 27-46. The clinical factors are lack of insight; negative attitudes; active symptoms of major mental illness; impulsivity; unresponsive to treatment. See *id.* at 49-58. The risk management factors are: plans lack feasibility; exposure to destabilizers; lack of personal support; noncompliance with remediation attempts; and stress. See *id.* at 61-70.

²¹⁶ See Kroner & Mills, *supra* note 191, at 474.

²¹⁷ See Dolan & Doyle, *supra* note 157, at 305.

²¹⁸ See M. GRANN, PERSONALITY DISORDER AND VIOLENT CRIMINALITY WITH SPECIAL REFERENCE TO PSYCHOPATHY AND RISK ASSESSMENT (1998).

²¹⁹ See J. Bonta et al., *The Prediction of Criminal and Violent Recidivism Among Mentally Disordered Offenders: A Meta-Analysis*, 123 PSCHOL. BULL. 123 (1998).

²²⁰ See Kroner & Mills, *supra* note 191, at 480 (citing studies).

c. VRAG

Of the three instruments addressed here, the most statistically sophisticated is the VRAG.²²¹ The goal of the VRAG study was to create an actuarial instrument capable of predicting which offenders would commit at least one violent act, given the opportunity.²²² “Violence” is measured by any new criminal charge (or act in subsequent institutionalization that would otherwise have resulted in a criminal charge) for a violent offense.²²³ Violence was defined as homicide, attempted homicide, kidnapping, forcible confinement, wounding, assault causing bodily harm, rape, armed robbery, and all sexual assaults involving physical contact.²²⁴ Opportunity was defined as being released into the community, a half-way house, or a minimum-security psychiatric hospital.²²⁵

A large number of independent variables (over fifty) were examined, and twelve were selected for the instrument (based on independent empirical data), and given “Pearson correlations” (a weighted number, each of which is added to yield a final score).²²⁶ These twelve variables had been independently reported by independent investigators.²²⁷ They consist of the following: Revised Psychopathy Checklist Score; Elementary School Maladjustment Score; meets DSM-III criteria for personality disorder; age at the time of the initial offense; separation from either parent under age sixteen (except for death); failure on prior conditional release; nonviolent offense history score; never married (or equivalent); DSM III criteria for schizophrenia; serious victim injury; alcohol abuse score; whether the victim of the index offense was female.²²⁸ Scoring using the instrument is accomplished by examining each of the factors and adding the scores of each. This total score is then placed within one of nine categories corresponding to incrementally greater rates of

²²¹ See *id.* at 482, tbl. 3 (finding ROC values for “major misconduct”—defined as inciting a riot, drug/alcohol use; refusing urinalysis; assaults; and refusing a direct order—of .575 for the PCL-R; .565 for the HCR-20, and .627 for the VRAG). Note that chance is defined as an ROC of .50, making none of these instruments fabulously predictive. Moreover, the outcomes measured are qualitatively different from the kind of violence that concerns death penalty juries.

²²² See QUINSEY, *supra* note 2, at 145.

²²³ See *id.* at 142.

²²⁴ See *id.*

²²⁵ See *id.* at 145.

²²⁶ See *id.* at 147; see also *supra* note 216.

²²⁷ See QUINSEY, *supra* note 2, at 145.

²²⁸ See *id.* at 147.

recurring violence.²²⁹

The instrument was tested on institutionalized offenders who had an opportunity to recommit over 81.5 months.²³⁰ The subjects of the study were 618 men institutionalized at a high security mental institution.²³¹ Of the 618 subjects studied, 191 (31%) committed new acts of violence.²³² If the scores of the subjects were divided into “high” and “low” risk groups, fifty-five percent of the “high risk” group committed new violent acts, compared to nineteen percent of the “low risk” group.²³³

Using sophisticated statistical analyses, the VRAG authors determined that they had achieved significant results.²³⁴ The authors acknowledge that the accuracy of the instrument depends on the base rate of violent recidivism in the population being studied, so that at very low base rates predictions would not be worthwhile.²³⁵ Independently, other investigators tested the instrument and found it superior to other instruments used to identify future misconduct.²³⁶

The authors of the VRAG attribute its accuracy to having measured the variables with reference to “detailed and complete psychosocial histories” gleaned from collateral sources (i.e., family members), descriptions of actual behavior rather than self-reporting, and “reliance on hypothetical internal psychological states” as its basis.²³⁷ Although they acknowledge that the instrument is not nearly as accurate as concurrent diagnostic tests²³⁸ or short-term weather predictions,²³⁹ the authors contend

²²⁹ See R. Mark Binderman, *Understanding VRAG: The Violence Risk Assessment Guide*, FORENSIC EXAMINER, Jan-Feb. 2001, at 28, 29 (using the example of an individual with a score of twenty-three, placing the individual in the category of subjects who had a 76% rate of recurring violence over seven years, and 98% more violent than the control group in the original study).

²³⁰ See QUINSEY, *supra* note 2, at 142.

²³¹ See *id.* at 145. Although these were mental patients, the authors noted that their characteristics were not very different from those of correctional institutions, and although they suffered from psychoses at a higher rate, that is a variable not likely to increase violent recidivism in offender populations. See *id.* (citing Bonta et al., *supra* note 219).

²³² Binderman, *supra* note 229, at 29.

²³³ WEBSTER ET AL., *supra* note 1, at 33 (computing risks of dangerousness).

²³⁴ See *infra* notes 313-16 and accompanying text.

²³⁵ See QUINSEY, *supra* note 2, at 152 (noting that “when the base rate is low enough, the optimal decision is to release everyone”).

²³⁶ See *id.* at 151.

²³⁷ *Id.* at 165 (observing that one reason for this may be that “the defining properties of psychopathy (lying, conning, and glib speech)” may increase measurement error during personal interviews).

²³⁸ The accuracy of concurrent diagnostic tests achieves an ROC of over .90, as compared with the VRAG’s .73. *Id.* at 168.

²³⁹ The authors insist that “the prediction of weather phenomena (where there is essentially zero outcome measurement error) probably cannot be accomplished in the prediction of violent recidivism.” *Id.* John Monahan and Henry Steadman similarly draw

that the instrument at least rivals the accuracy of predicting violent storms.²⁴⁰

IV. ASSESSING SCIENTIFIC VALIDITY OF ACTUARIAL PREDICTIONS

The scientific validity of actuarial predictions depends upon probabilistic assessment of data, independent verification of procedures, and concern over false positives.²⁴¹ Scientists recognize that the precise prediction of future behavior is impossible—whether the subject is the weather or human violence, the most that can be predicted is its probability. The scientific validity of such predictions requires evaluation of probabilistic reasoning, an evaluation that focuses on falsifiability, interdisciplinarity, and rationality, emphasizing the explanatory power of a proposed hypothesis.²⁴²

Scientific predictions are a form of risk analysis, a statistical methodology commonly used by epidemiologists, toxicologists, the Environmental Protection Agency, and the insurance industry, among others—but is it scientifically valid for predicting human behavior rather than disease? The answer to that question requires an evaluation not only of the mathematical logic, but also of the underlying theory. Statistical analysis provides an important tool for examining whether theories correspond with observation.²⁴³ However, one cannot expect statistics to provide an answer about any particular individual.²⁴⁴ The most that can be said from even the best statistical analyses is that someone falls within a group that has a certain statistical propensity for

an explicit analogy to weather reports in advocating the use of actuarial instruments and probabilistic communication for predicting violence. See John Monahan & Henry Steadman, *Violent Storms and Violent People: How Meteorology Can Inform Risk Communication in Mental Health Law*, AM. PSYCHOLOGIST, Sept. 1996, at 931.

²⁴⁰ See QUINSEY, *supra* note 2, at 168.

²⁴¹ See generally Beecher-Monas, *supra* note 178.

²⁴² See *id.* at 1579.

²⁴³ See Kenneth J. Rothman, *Significance Questing*, 105 ANNALS INTERNAL MED. 445, 445 (1986) (editorial).

²⁴⁴ As epidemiologists studying the statistical incidence of disease have discovered, one simply cannot predict the probability of disease in any given case. See Sander Greenland & Jerome Robins, *Epidemiology, Justice, and the Probability of Causation*, 40 JURIMETRICS 321, 328 (2000) (explaining that “when an exposure is known to be harmful in some cases, available data from epidemiology and biology are simply incapable of telling us whether a given case was ‘more probably than not’ harmed by exposure”). The most that can be said is that exposure caused a certain statistical increase of disease over background levels.

violence.²⁴⁵ Moreover, statistical measurements must be understood within the context of the system being studied.²⁴⁶ Here, the context is human behavior, a quintessentially complex phenomenon. Complexity theory recognizes the continuous interaction of individuals with the whole system of which they are a part, and who respond to both random events and dynamic phenomena.²⁴⁷

The framework for this analysis is taken from an earlier article, in which I argued that in order to meet *Daubert* criteria a judge must be able to do five things: (1) identify and examine the proffered theory and hypothesis for their power to explain the data; (2) examine the data that supports (and undermines) the expert's theory; (3) use supportable assumptions to fill the inevitable gaps between data and theory; (4) examine the methodology; and (5) engage in probabilistic assessment of the link between the data and the hypothesis.²⁴⁸ This framework seeks to provide a structured reasoning process for examining scientific validity that goes beyond—while satisfying—the sketchy *Daubert* guidelines of testability, peer review, and publication, existence of protocols and error rates, and general acceptance.²⁴⁹ In each section of the analysis (i.e. the explanatory power of the theory/hypothesis), satisfaction of the inquiry also satisfies one or more of the *Daubert* factors, although the underlying inquiry goes to the heart of what I believe the *Daubert* court was aiming for.²⁵⁰

²⁴⁵ The concept of probability I am referring to is the idea of long-run relative frequency. See THEODORE COLTON, STATISTICS IN MEDICINE 63 (1974) (defining the “probability of an event [as] the event’s long-run relative frequency in repeated trials under similar conditions”). That is, the probability of recurring violence is its relative frequency of occurrence—or the proportion of times the event occurs—in a large number of trials repeated under virtually identical conditions. *Id.* “Virtually identical conditions” are hard to come by in observational studies of humans (such as the studies underlying the violence risk assessment tools). This is a pervasive problem for human studies, but one that does not necessarily undermine their validity. See Beecher-Monas, *supra* note 178, at 1604-07 (discussing scientific validity of human studies).

²⁴⁶ See COLTON, *supra* note 245, at 117, 304 (explaining that although a result may be statistically significant, it may still be medically meaningless, and cautioning against permitting statistical analyses to generate hypotheses).

²⁴⁷ See *id.* at 20 (noting the “continuous conversation between parts and wholes”).

²⁴⁸ See Beecher-Monas, *supra* note 178, at 1571.

²⁴⁹ See *id.* at 1567.

²⁵⁰ See generally *id.*

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

1. Statistical Theory

The theory underlying each of the actuarial instruments used in predicting future violence is first that violent behavior is statistically correlated²⁵¹ with specific factors either in peoples' past behavior (a pattern of past violence, for example), their circumstances (poverty, for example), their attitudes toward others (failure to marry or form equivalent relationship), their medical and psychiatric history (age when problems began, and any injuries to the brain), and substance abuse (alcohol or drugs); and second, that these factors in combination can be used to probabilistically assess a level of risk for the future. The variables considered for the instruments were drawn from empirical studies showing a statistical association²⁵² with violent behavior.

In constructing the VRAG, for example, an offender population was divided into two groups, one that had committed a further act of violence and one that had not.²⁵³ The overall base rate for re-offending was thirty-one percent.²⁵⁴ Violent offenders who had married had a "violent recidivism rate" (that is they re-offended at a rate) of twenty-one percent, while those who had not

²⁵¹ See GOULD, *supra* note 99, at 269, explaining that "[c]orrelation assesses the tendency of one measure to vary in concert with another." Correlation is measured—for linear relationships—by Pearson's product moment correlation coefficient, which ranges from +1 for perfect positive correlation, 0 for no correlation, to -1 for perfect negative correlation. *See id.* at 270.

²⁵² Association means that there is a statistically significant correlation of a particular factor with violent behavior. Statistical significance is set by convention at a level of significance, or p-value of .05 (which corresponds to a confidence level of 95 percent). The object of statistical significance tests is to keep the scientist from asserting a positive effect when the effect may actually be due to chance. *See* David Ozenhoff & Leslie I. Bodin, *Truth & Consequences: Health Agency Responses to Environmental Health Problems*, 12 SCI. TECH. & HUM. VALUES 70, 73-74 (1987). If the p-value is .01, the evidence is said to be highly statistically significant. *See* Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 JURIMETRICS J. 1, 22 (1995). "By rejecting a hypothesis only when the test is statistically significant, we have placed an upper bound, 5%, on the chance of rejecting a true hypothesis." *Id.* Another way of explaining this is that it describes the probability that the procedure produced the observed effect by chance. *See id.* If the test is not statistically significant, it may either be because the results were due to chance or because the test lacked the power to discern a difference between the null hypothesis and the proposed effect. *See id.* Power increases with the size of the study and with the degree of difference from the null hypothesis (the more extreme the alternatives, the better the power). *See id.*

²⁵³ *See* QUINSEY, *supra* note 2, at 75.

²⁵⁴ *Id.* at 147.

married re-offended at a rate of thirty-eight percent.²⁵⁵ This is not a theory that failure to marry causes violence; as a technique, it is more like diagnosing illness, where a “convergence of signs, symptoms, outcome, and patterns of familial aggregation” are examined with only the sketchiest of underlying causal theories.²⁵⁶ The most that can be said is that marriage and a decreased propensity for violence are statistically associated; they are found together in a significant percentage of the population being studied.

These factors are not “causes” of violence.²⁵⁷ Not everyone who is poor is violent, for example, but the level of violence in poor areas is higher than in wealthy areas. Therefore, poverty is associated with violence, although it cannot be said to “cause” violence. Is this testable? Certainly—identify the null hypothesis (that the factor is not statistically correlated with a particular outcome—violence, in this case), define your terms, run your numbers on a sufficiently large population enough times, and you have tested your hypothesis. But what does it mean if you find a statistically significant correlation?²⁵⁸

Statistical theory is based on the idea that there is an underlying structure in large matrices of data.²⁵⁹ The problem with this theory is that, although the mathematics are unimpeachable, neither the fact that there is a correlation nor the strength of the correlation says anything about the nature of the cause.²⁶⁰ For example, as the number of my publications increases, the price of diamonds has also increased, and so has the distance between the Earth and Halley’s Comet.²⁶¹ There may be a strong positive correlation among these factors.²⁶² Are they causally related (even

²⁵⁵ *Id.*

²⁵⁶ Nancy C. Andreasen, *Linking Mind and Brain in the Study of Mental illnesses: A Project for a Scientific Psychopathology*, 275 *Sci.* 1586, 1587 (1997) (discussing conceptual issues in scientific psychopathology).

²⁵⁷ See Stephen F. Lanes, *Error and Uncertainty in Causal Inference*, in *CAUSAL INFERENCE* 173, 182 (Kenneth J. Rothman ed., 1988). “The uncertainty in causal inference is attributable to the fact that we cannot establish that an association is valid.” *Id.* at 185. An unidentifiable error may exist and it may cause the observation. See *id.* The most that can be expected of strength of association and the level of statistical significance is that they affect subjective beliefs.

²⁵⁸ Meaning, or rationality, is one of the facets of falsifiability. It is a way of comparing ideas, and consists of openness to critique, testing the theory against other theories in terms of coherence and empirical content and explanatory power. See Karl Popper, *Normal Science and its Dangers*, in *CRITICISM AND THE GROWTH OF KNOWLEDGE* 51, 57 (Imre Lakatos & Alan Musgrave eds., 1970).

²⁵⁹ See GOULD, *supra* note 99, at 268.

²⁶⁰ *Id.* at 273.

²⁶¹ See *id.* at 272 (explaining that even the most dedicated astrologer would not discern causality in most of these relationships).

²⁶² As Webster explain with regard to correlations, “some findings may be very highly

in a probabilistic sense)? Obviously not. The majority of statistically correlated factors are merely coincidental. Thus, the choice of factors for analysis implies some underlying causal theory even if it is unstated (or undeveloped).²⁶³

Moreover, even if there is a causal relationship, correlations can tell us little about the nature of the cause.²⁶⁴ The positive statistical correlation between poverty and violence does not show anything about which one may be a cause of the other or what the causal nature might be.²⁶⁵ For example, if a propensity for violence correlates with socioeconomic situation, is that because poverty breeds despair, or because poverty causes poor childhood nutrition, which affects the growing brain (and therefore decisionmaking), or because there is a high level of lead paint (linked to neural deficits) in poor neighborhoods? Or is it because violent people have trouble relating to others and so they become poor through an inability to hold a job? Or is the correlation due to something else entirely? Thus, the question underlying the choice of factors used in each of the instruments must be: Why this factor and not others?²⁶⁶

Notably, none of the actuarial instruments claims a causal connection, although the PCL-R comes close.²⁶⁷ Only the PCL-R advances an express theory²⁶⁸ for why it chose any particular

statistically significant but . . . they may be extremely unimportant or uninteresting.” WEBSTER ET AL., *supra* note 1, at 15 n.11.

²⁶³ See MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS* 379 (2d ed. 2001) (explaining that a labor economist might use labor market theory to select certain types of experience and education as explanatory factors in a wage equation but noting that this ideal is seldom met in practice, either because theory is not complete enough to dictate the choices, or because data for theoretically perfect factors are not available and surrogates must be used instead).

²⁶⁴ See GOULD, *supra* note 99, at 273.

²⁶⁵ See, e.g., Eric Silver et al., *Assessing Violence Risk Among Discharged Psychiatric Patients: Toward an Ecological Approach*, 23 L. & HUM. BEHAV. 237, 250 (1999) (observing that “concentrated poverty in the neighborhoods where patients resided after discharge was significantly related to the overall amount of violence they committed”). This correlation says nothing about cause. Similarly, just because war has been an enduring fact of human history does not mean that violence is an innate trait of human beings; rather, the potential for aggression may be an expression of some other underlying biological principles that anticipate peaceful interactions in other environments. See GOULD, *supra* note 99, at 360 (critiquing the assertions of E.O. Wilson that aggression is innate in human beings). In other words, there is a wide range of human behaviors that are different expressions of biological potentials in various environments. See *id.* at 360.

²⁶⁶ See Naomi Oreskes et al., *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences*, 263 SCI. 641, 641 (1994) (using the example of numerical simulation models in the earth sciences, and noting that “the observation and measurement of both independent and dependent variables are laden with inferences and assumptions”).

²⁶⁷ See Hare, *supra* note 192, at 99 (claiming that “[p]sychopathy is the prime criminogenic personality trait”).

²⁶⁸ See *id.* at 100, 103 (claiming that psychopathy is a distinct type of personality

factors.²⁶⁹ If the PCL-R is flawed in its reasoning, however, it affects all three of the instruments, because both of the other instruments use the PCL-R as one of their factors.

The underlying theory of any actuarial violence risk instrument is that if you combine a critical number of risk factors, each of which has been demonstrated to have a statistically significant correlation with a recurrence of violent behavior, as long as you choose the factors carefully (by controlling for inter-relationships among the various factors—such as previous criminal charges and prior convictions)²⁷⁰ you can predict the probability of recurring violent conduct. The theory underlying the creation of such an instrument is that patterns of behavior in populations can predict the probability that an individual's behavior will fall within that range. Is this assertion falsifiable? At least in theory it is, although in practice (because of the large sample sizes needed and the length of time necessary to assess recurrence) it may be difficult. A more fundamental problem is whether prediction is ever possible in biological systems.

exhibiting “a constellation of affective, interpersonal, and behavioral characteristics” and that they “are responsible for a disproportionate amount of the serious crime in most societies”). The problem with this assertion, among other things, is that not only violent individuals share these traits, but so do those who do not break the law, together with others whose crimes “run the gamut from petty theft and fraud to cold-blooded violence.” *Id.* at 103. Although Hare claims a base level of psychopathy at less than one percent of the general population, he also claims that “they comprise a significant proportion of our prison population.” *Id.* at 104. Surprisingly, he further claims that “it is not uncommon for psychopaths to emerge as leaders and ‘patriots’” as well as “high pressure salesmen and stock promoters,” “radical political activists,” people “active in the business and corporate world,” “unethical lawyers and doctors” and a panoply of “serial killers and rapists, drug dealers, pimps, spouse and child abusers, swindlers and con men,” and just about every other possible human occupation (except psychologists). *Id.* Wait a minute: wasn't this instrument supposed to predict violence? Except for the “serial killers” and “terrorists” in Hare's panoply, those described are not committing acts of physical violence. So just what is being predicted? Nonetheless, Hare asserts that the PCL-R is a “robust predictor of violence in criminals,” and a number of researchers agree. See Dolan & Doyle, *supra* note 157, at 305 (citing studies); see also WEBSTER ET AL., *supra* note 1, at 12 (citing a study in which the PCL-R score “alone performed as well as, or better than, the best combination of 16 other variables selected by multiple regression”). Despite the accolades of a number of researchers, Hare's theory is highly questionable when examined in light of Popper's criteria for falsifiability. See Popper, *supra* note 258, at 57. One reason for the success of the PCL-R may be its reliance on factors associated with childhood aggression that may have a biological basis quite different from that asserted by Hare.

²⁶⁹ The VRAG factors, for example, were chosen because of their “well-established empirical relationship with violent criminality.” QUINSEY, *supra* note 2, at 143. The authors of the instrument acknowledge that improvements to their instrument are possible by including “physiological measures (e.g., serotonin levels) or situational variables (e.g., postdischarge social supports)” but that the research to show whether these variables would improve the predictive accuracy of the VRAG has not been done. *Id.* at 176-77.

²⁷⁰ See *id.* at 146.

2. Complexity Theory and the Problem of Prediction

Predictions in complex systems are always highly contingent, and human behavior is an emblematic complex system. “A complex adaptive system is a collection of individual agents with freedom to act in ways that are not always totally predictable, and whose actions are interconnected so that one agent’s actions changes the context for other agents.”²⁷¹ The principles of complexity theory have largely replaced deterministic notions of causation in biology, and some aspects of physics as well.²⁷² Complexity theory explains that we, as individuals, are interacting parts of a complex world, we have numerous interactions with our environment, and the instigator of our actions, the brain itself, is a complex organ.²⁷³

Interactions of complex systems simply cannot be predicted far in advance with any accuracy. We cannot predict the occurrence of an idea or the behavior that results from it any more than we can predict next week’s weather.²⁷⁴ We can discern regular patterns of weather which recur under particular conditions, but the weather will change in unpredictable ways when any of the underlying conditions change.²⁷⁵ The problem is two-fold: first, small errors in determining the initial conditions (i.e., the predictors) may yield large errors in calculating expected outcomes; and second, even when the properties of the individual components are understood, the behavior of a system with many

²⁷¹ Paul E. Plsek & Trisha Greenhalgh, *The Challenge of Complexity in Health Care*, 323 BRIT. J. MED. 625 (2001) (giving as examples the immune system, a colony of termites, the financial market, and “just about any collection of humans”).

²⁷² See ILYA PRIGOGINE, *THE END OF CERTAINTY: TIME, CHAOS AND THE NEW LAWS OF NATURE* 4-5 (1997) (explaining that while “[c]lassical science emphasized order and stability; now, in contrast, we see fluctuations, instability, multiple choices, and limited predictability at all levels of observation. . . . [so that] we are now able to include probabilities in the formulation of the basic laws of physics”).

²⁷³ See Mark D. Albertson, *Can Violence Be Predicted? Future Dangerousness: The Testimony of Experts in Capital Cases*, 3 CRIM. JUSTICE 18, 45 (1989) (explaining that a “person-focused” assessment . . . is extremely inaccurate because people do not live in vacuums” and research emphasizes “the importance of situational and environmental influences on behavior”).

²⁷⁴ We cannot predict the weather because it is a “classic case of chaotic behavior.” RICARD V. SOLE & BRIAN GOODWIN, *SIGNS OF LIFE: HOW COMPLEXITY PERVADES BIOLOGY* 3, 9 (2000). That is because “small errors in initial conditions give rise to very large errors in calculating expected outcomes.” *Id.* at 12.

²⁷⁵ Complexity theory, the study of nonlinear systems (like the weather), involves both the study of chaos, with sensitivity to initial conditions that makes dynamics unpredictable, and emergent properties, in which observers are generally unable to predict the behavior of nonlinear systems from their parts and interactions. *See id.* at 20.

interacting components is inherently unpredictable.²⁷⁶

Complexity theory explains the notorious difficulty of predicting violent storms.²⁷⁷ As one judge noted, "A weather forecast is a classic example of a prediction of indeterminate reliability, and a place peculiarly open to debatable decisions Weather predictions fail on frequent occasions."²⁷⁸ Why should this be so? The weather is a classic case of chaotic behavior (where small errors in measurement of initial conditions give rise to grossly inaccurate predictions) and emergent phenomena (where many components interact).²⁷⁹ Weather is due to "the behavior of the gases that make up the earth's atmosphere under the influence of the rotation of the earth and the sun's radiant energy."²⁸⁰ Although the atmosphere obeys the equations of fluid dynamics and should therefore be determinable, it also is governed by the nonlinear dynamics of the Lorenz attractor, a mathematical description of the trajectories of motion of atmospheric flow in three dimensions.²⁸¹ As a result, the weather patterns display both order and chaos. What this means in terms of predictability is that weather predictions are fairly accurate for the next day, but fall off rapidly for three-day forecasts, and become highly chaotic after six days.²⁸²

The importance of measuring relevant information is stressed in complexity theory. In weather prediction, for example, "over 10,000 land-based stations and hundreds of ships collect weather information daily at six-hour intervals."²⁸³ Feedback about predictions is crucial. In addition to the weather stations, there are satellites, balloons, aircraft, and human spotters making daily observations, and funneling them to one of several meteorological centers.²⁸⁴ These centers generate regional reports, which are then adapted to local conditions.²⁸⁵ A huge amount of information is generated and analyzed at frequent intervals, something unlikely to be achieved in human behavior.

²⁷⁶ See *id.* at 13.

²⁷⁷ Although the assertion is made that weather forecasters are "exemplary risk assessors," the error rate is actually quite high, particularly over the long-term and for rare and severe weather events, and especially when the amount of data used is taken into consideration. See Monahan & Steadman, *supra* note 239, at 931.

²⁷⁸ *Brown v. United States*, 790 F.2d 199, 204 (1st Cir. 1986).

²⁷⁹ See SOLE & GOODWIN, *supra* note 274, at 9.

²⁸⁰ *Id.*

²⁸¹ See *id.* at 10.

²⁸² See R. Kerr, *Official Forecasts Pushed Out to a Year Ahead*, 266 SCI. 1940, 1940 (1994) (stating that weather predictions are "swamped by chaos beyond six days or so").

²⁸³ Monahan & Steadman, *supra* note 239, at 933.

²⁸⁴ See *id.*

²⁸⁵ See *id.*

This illustrates an important failing of the actuarial future dangerousness instruments. None of the actuarial instruments incorporates any of the biological information about violence. Given what we know about the importance of measuring initial conditions in complexity theory, this is a huge gap. All the available actuarial instruments, even the best, would be more predictive if they also assessed biological information about the brain.²⁸⁶ Those factors would include “physiological measures and assessments of neurocognitive function and how individuals process emotional information.”²⁸⁷

Moreover, even using all available data, there are limits to prediction.²⁸⁸ There are three reasons for these limits. First, the human brain is the premier example of nonlinearity: that is, there is no predictable relationship between cause and effect. The brain itself is composed of multiple interacting and self-regulating physiological systems including biochemical and neuroendocrine feedback loops, which influence human behavior partly through an internal set of responses and partly through adaptive responses to new stimuli from the environment, forming a web of interacting systems that are dynamic and fluid.²⁸⁹ Second, an individual’s conduct results from both internal stimuli and environmental stimuli, including a web of relationships affecting beliefs,

²⁸⁶ See Dolan & Doyle, *supra* note 157, at 309 (urging incorporation of biological factors into actuarial instruments).

²⁸⁷ *Id.*

²⁸⁸ See Robinson, *supra* note 142, at 1450.

²⁸⁹ See SOLE & GOODWIN, *supra* note 274, at 1. As an article in a noted scientific journal explained:

- The human body is composed of multiple interacting and self regulating physiological systems including biochemical and neuroendocrine feedback loops.
- The behaviour of any individual is determined partly by an internal set of rules based on past experience and partly by unique and adaptive responses to new stimuli from the environment.
- The web of relationships in which individuals exist contains many varied and powerful determinants of their beliefs, expectations, and behaviour.
- Individuals and their immediate social relationships are further embedded within wider social, political, and cultural systems which can influence outcomes in entirely novel and unpredictable ways.
- All these interacting systems are dynamic and fluid.
- A small change in one part of this web of interacting systems may lead to a much larger change in another part through amplification effects.

For all these reasons neither illness nor human behaviour is predictable and neither can safely be “modelled” in a simple cause and effect system. The human body is not a machine and its malfunctioning cannot be adequately analysed by breaking the system down into its component parts and considering each in isolation.

Tim Wilson & Tim Holt, *Complexity and Clinical Care*, 323 BRIT. J. MED. 685 (2001).

expectations, and behavior.²⁹⁰ Third, individuals and their immediate social relationships are further embedded within wider social, political, and cultural systems that are continuously interacting; nature, nurture, and notions of free will all interact in a way that can only be considered probabilistic. The equilibrium of each of these systems can be radically altered by seemingly inconsequential stimuli.

Thus, no matter how accurate the instrument, it will be impossible to predict a particular individual's behavior.²⁹¹ The most that can be said is a probabilistic statement for the group within which the individual fits. Why, then, would anyone propose that actuarial instruments be used for predicting violence? If the best that can be achieved is the contingent reliability of weather predictions, is that reliable enough to sentence someone to death? The best that can be said is that actuarial instruments may improve on the woeful inadequacy of clinical predictions.²⁹²

B. *What Data Supports or Undermines the Theory?*

The VRAG, the PCL-R, and the HC-20 have all been published and peer-reviewed in the psychological and criminology communities, thus satisfying the *Daubert* prongs of peer review and publication. They have been, overall, well-received, thus meeting general acceptance. But the important question in each case is whether the theory has factual support, or whether there is data that significantly undermines it.²⁹³

1. Choice of Variables

The VRAG study considered a large number of variables that had been shown by other researchers to have a demonstrated correlation with violence, and performed separate stepwise

²⁹⁰ See generally DAMASIO, *supra* note 117 (discussing the neural underpinnings of reason, emotion, and the complex, interactive systems of the brain, which in turn interact with systems in the rest of the body, the environment, other individuals, and culture).

²⁹¹ See Plsek & Greenhalgh, *supra* note 271, at 625.

²⁹² See WEBSTER ET AL., *supra* note 1, at 20 (noting that courts and legislatures continue to demand these predictions, although "three decades of research has failed to produce an accurate scheme for predicting violence" and outlining a "scheme for prediction which we hope will offer better accuracy").

²⁹³ See Beecher-Monas, *supra* note 178, at 1595 (urging an interdisciplinary analysis that examines all the available information in light of three questions: "What range of facts deserves investigation? What is the proper way to investigate them? And what do the results of the investigation mean?").

discriminant analyses in order to identify variables in various categories that had the greatest predictive power.²⁹⁴ These categories included childhood history, adult adjustment, and the circumstances of the initial violent offense.²⁹⁵ From these stepwise regressions, the researchers chose twelve variables.²⁹⁶ When combined, these twelve variables yielded an overall correlation (weighted to take into account the relative importance of each) of 0.45.²⁹⁷ This is not a high degree of correlation, but it represents an improvement over most previous actuarial violence predictions, the best of which managed a correlation of 0.40.²⁹⁸ If the most strongly correlated variables—the PCL-R score, elementary school maladjustment, and DSM III diagnosis of personality disorder—were omitted from the instrument, the overall correlation of the VRAG fell to 0.36, showing that the remaining variables demonstrate some, though less, predictive power.²⁹⁹ These low (though better than prior instruments’) correlations indicate that these instruments all have considerable unexplained variance.³⁰⁰

Although the VRAG includes the PCL-R, the authors also critique it. They found that the correlation between violent recidivism and the PCL-R alone is only .33, and that adding the other eleven VRAG variables increased the correlation to .45.³⁰¹ They contend that “the entire PCL-R can be replaced by variables pertaining only to antisocial childhood behavior imply[ing] that PCL-R factor 1 items reflecting apparently adult personality (e.g., glibness, grandiosity, lying, conning, remorselessness, shallowness, callousness) do little or nothing to reduce uncertainty about the likelihood of violent recidivism.”³⁰² Thus, the link is much stronger between childhood aggression and adult violence than between any of the so-called personality traits. Moreover, the VRAG authors criticized PCL-R’s reliance on clinical judgment, so they replaced clinical assessment of the PCL-R with factors that could be obtained from the subject’s history.

²⁹⁴ See WEBSTER ET AL., *supra* note 1, at 31.

²⁹⁵ *See id.*

²⁹⁶ *See id.*

²⁹⁷ A perfect correlation would be 1.0. SPATZ, *supra* note 91, at 80.

²⁹⁸ WEBSTER ET AL., *supra* note 1, at 36-37. As a point of comparison, LSAT scores, which are considered to be predictive enough of law school performance to base admission on them, have a correlation coefficient of about 0.60. FINKELSTEIN & LEVIN, *supra* note 263, at 30.

²⁹⁹ See WEBSTER ET AL., *supra* note 1, at 36-37.

³⁰⁰ See Kroner & Mills, *supra* note 191, at 486 (noting that this may be due to “inadequacies in conceptualization and, of equal importance, instrumentation”).

³⁰¹ WEBSTER ET AL., *supra* note 1, at 163 (noting that the increase under the ROC method was from .65 to .75).

³⁰² *Id.* at 168.

2. Inter-rater consistency

The consistency with which different people assess the risk of violence in a particular person is an important consideration, especially in litigation. Although the VRAG has been carefully structured in an attempt to minimize cognitive biases, all of the actuarial instruments require human input and at least a degree of subjective decisionmaking. Some degree of variability in scoring can therefore be expected among different raters. The VRAG inter-rater agreement in scoring was found to be 0.90, showing a high degree of correlation.³⁰³ The PCL-R inter-rater agreements “typically exceed 0.80.”³⁰⁴ Inter-rater reliability is a concern, because if two raters disagree about a prediction, at least one of them will be wrong.³⁰⁵ Inter-rater agreement however, is no guarantee of validity; when all raters agree, the instrument may still be invalid—the raters could all be wrong.³⁰⁶

3. Measuring Predictive Accuracy

The preferred method for evaluating the prediction accuracy of each of the instruments discussed is the statistical analysis known as receiver operator characteristics (ROC).³⁰⁷ This is also the type of statistical analysis preferred for comparing instruments to each other.³⁰⁸ In this type of analysis, sensitivity³⁰⁹ and specificity³¹⁰ are taken into effect to yield an area under the

³⁰³ WEBSTER ET AL., *supra* note 1, at 41 n.4. Correlations can be either positive (+) or negative (-); for ease of reference, the correlations referred to in this Article are positive unless otherwise noted.

³⁰⁴ Hare, *supra* note 192, at 101.

³⁰⁵ See Christopher Slobogin, *Doubts About Daubert: Psychiatric Anecdotes as a Case Study*, 57 WASH. & LEE L. REV. 919, 921 (2000) (noting that “validity is suspect” when two raters fail to agree).

³⁰⁶ See *id.*

³⁰⁷ For a discussion of the use of ROC analysis in predicting violence in involuntary hospitalization proceedings, see Douglas Mossman, *Dangerousness Decisions: An Essay on the Mathematics of Clinical Violence Prediction and Involuntary Hospitalization*, 2 U. CHI. L. SCH. ROUNDTABLE 95, 115 (1995) (explaining that “the relative performance of diagnostic tests can be evaluated using ROC graphs”).

³⁰⁸ See Dolan & Doyle, *supra* note 157, at 304.

³⁰⁹ Sensitivity in the violence prediction instruments is the rate of violent outcomes when risk had been predicted. See *id.* at 309.

³¹⁰ For the violence prediction instruments, this would be the rate of nonviolent outcomes when no risk had been predicted. See *id.* Notably, there are trade-offs in sensitivity and specificity, and ROC analysis provides a method for assessing accuracy independent of biases favoring sensitivity or specificity and also distinct from the prevalence of violence. See Mossman, *supra* note 307, at 126 (describing ROC method to

measured curve of scale scores.³¹¹ A random prediction (chance) would yield an ROC value of 0.5; ROC values range from 0-1, with 1 representing nearly perfect accuracy.³¹² An ROC of .75 is considered to be a large effect size.³¹³

To measure the accuracy of the VRAG, the authors used a statistical analysis known as the Relative Operating Characteristic (ROC).³¹⁴ This method of analysis was chosen because of the low base rate in the populations studied.³¹⁵ A test that is no better than chance would have an ROC of .50; the VRAG ROC was .76, which means that “if an offender were drawn randomly from each of the recidivist and nonrecidivist groups, there was a probability of .76 that the recidivist had the higher score on the VRAG.”³¹⁶ This is a statistically significant result, comparable to ROC scores for predictions in meteorology and medical imaging.³¹⁷ Studies on the PCL-R have shown an ROC value of approximately .72.³¹⁸ A study of post-release violent offenders comparing the VRAG with the PCL-R and the HCR-20 showed an ROC value of .596, .618, and .562, respectively, for subsequent violent convictions over a mean release time of 790 days, but found no statistically significant difference among the instruments.³¹⁹ Note that with chance defined as an ROC of .50, none of these instruments is fabulously predictive.

C. *What Assumptions are Used to Fill the Gaps and Are They Supportable?*³²⁰

Probabilistic assessments inevitably employ assumptions,³²¹

predict accuracy).

³¹¹ See *id.* at 304-05 (describing the statistical analysis).

³¹² See *id.* at 304.

³¹³ See *id.* at 305.

³¹⁴ See Binderman, *supra* note 229, at 29.

³¹⁵ See *id.*

³¹⁶ QUINSEY, *supra* note 2, at 148.

³¹⁷ See Binderman, *supra* note 229, at 29.

³¹⁸ See Dolan & Doyle, *supra* note 157, at 305 (citing GRANN, *supra* note 218).

³¹⁹ Kroner & Mills, *supra* note 191, at 479, 483. These same analysts found ROC values for “major misconduct”—defined as inciting a riot, drug/alcohol use; refusing urinalysis; assaults; and refusing a direct order—of .575 for the PCL-R; .565 for the HCR-20; and .627 for the VRAG. *Id.* at 482, tbl. 3. These outcomes however, are qualitatively different from the kind of violence that concerns death penalty juries.

³²⁰ The inquiry into supportable assumptions corresponds to some degree to *Daubert's* General Acceptance factor, in that assumptions that are not commonly held will be controversial.

³²¹ All knowledge is incomplete; one simply cannot calculate “the probability that a theory is true.” Stephen F. Lanes, *The Logic of Causal Inference*, in CAUSAL INFERENCE 64-65 (Kenneth J. Rothman ed., 1988). As a result, assumptions to fill the gaps in

and the actuarial instruments used in predicting dangerousness are no different in this respect. Whether these assumptions are supportable depends in large part on consensus among the scientific community and the empirical basis for them.³²² As the authors of the VRAG acknowledge, the more assumptions that are made about the data, the less predictive the resulting instrument.³²³

One major assumption of all the instruments is that the population that was the subject of the underlying studies can be generalized to the people on whom the instrument will be used to predict future dangerousness. Each of the instruments was developed based on studies of men incarcerated in psychiatric hospitals or prisons. That does not seem too great a stretch, because the patients studied had all been admitted for crimes of violence, a population similar to convicted defendants facing sentencing proceedings.

The necessary assumptions for regression analysis include the assumption that the model has been correctly specified; that any measurement errors are random and independent; that measurement errors are independent of the corresponding observations for each of the model's explanatory variables; and that no explanatory variable is perfectly correlated with a combination of other variables.³²⁴ Moreover, the assumption of non-linearity and interactive models (as discussed above in the complexity section) would improve accuracy.³²⁵ Other methods, such as neural networks and entropy minimax, have been used to discover predictive relationships.³²⁶

knowledge are inevitable. Beecher-Monas, *supra* note 178, at 1583.

³²² See IRME LAKATOS, *THE METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES* 49 (1978).

³²³ See QUINSEY, *supra* note 2, at 146.

³²⁴ See Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, in *REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* 181, 213 (Federal Judicial Center 2d ed. 2000).

³²⁵ *But see* QUINSEY, *supra* note 2, at 146 (arguing that these improvements would make little difference, given the possibility of measurement error and the level of accuracy already achieved). An example of nonlinearity is the correlation of arousal and performance where, although there is a strong correlation between performance and arousal, the relationship is a curved line (nonlinear), showing low performance at both high and low arousal and optimal performance in the mid-range. See SPATZ, *supra* note 91, at 96 (graphing the relationship between arousal and efficiency of performance). A linear relationship between cause and effect can be demonstrated by turning the dimmer switch on a light; the illumination varies in direct proportion to the amount the dimmer switch is turned. See SOLE & GOODWIN, *supra* note 274, at 1-2.

³²⁶ See QUINSEY, *supra* note 2, at 146, 168 (citing studies and acknowledging that other kinds of studies might be an improvement, but arguing that the VRAG's accuracy rivals that for predicting violent storms).

D. Methodology

The hypothesis underlying the VRAG was tested through a retrospective study of a population of 618 released offenders over a period of seven years.³²⁷ The effect of each factor on overall risk was scored, and the percentage of people who had re-offended after the release was correlated with the score.³²⁸ The higher the score, the more violence was committed by that particular group.

The VRAG study used a large sample (618 people), permitting a statistical analysis. Its definitions of violent conduct appear to be the kinds of behavior capital jurors might worry about: killing, attempted killing, kidnapping, forcible confinement, wounding, assault causing bodily harm, pointing a firearm, rape, and sexual assaults involving robbery. This definition probably corresponds to the types of behaviors a jury would be concerned about in deciding the issue of future dangerousness.

In constructing the VRAG, variables were selected on the basis of having a significant correlation with recurring violence.³²⁹ If any variables were highly correlated with each other, the one with the lowest correlation with recurring violence was dropped.³³⁰ Least-squares stepwise multiple regression analysis was employed “to select variables that added independently to the prediction of violent recidivism.”³³¹

The particular statistical tool used by the actuarial instruments is multiple regression, a method for probing the relationship of two or more variables.³³² Multiple regression takes a variable that must be explained (recurring violence, in this instance) and examines its association with other variables such as

³²⁷ See WEBSTER ET AL., *supra* note 1, at 30.

³²⁸ This was the testing methodology that was used to construct the VRAG. See QUINSEY, *supra* note 2, at 141-51.

³²⁹ See *id.* at 145.

³³⁰ See *id.* at 146.

³³¹ *Id.* Two events, A and B, are independent when $P(A/B) = P(A)$. As Suppes explains, “the concept of independence is one of the most profound and fundamental ideas, not only of probability theory but of science in general.” PATRICK SUPPES, A PROBABILISTIC THEORY OF CAUSALITY 8 (1970). The concept of independence asserts that discrete biological or physical events, like coin tosses, have no memory of time and place. See Lynn A. Steen, *The Science of Patterns*, 240 SCI. 611 (1988). In other words, knowing whether a coin toss yielded heads or tails last time will not help predict what the next coin toss will yield. See FREDERICK MOSTELLER ET AL., PROBABILITY WITH STATISTICAL APPLICATIONS 8 (2d ed. 1970).

Independence is when “two variables are not correlated with each other in the population.” Rubinfeld, *supra* note 324, at 179.

³³² See *id.* at 181.

poverty.³³³ Multiple regression is a way of relating one variable to the values of other explanatory variables in order to predict the value of one variable using the values of others.³³⁴ It is frequently employed in antitrust litigation, anti-discrimination class actions, and market manipulation cases.³³⁵

Least-squares regression was the method used in constructing the VRAG. Least-squares is a well-regarded technique for estimating the underlying parameters.³³⁶ Least-squares is a method of reducing error when making predictions from a linear relationship.³³⁷

To construct the VRAG, separate regression analyses were run on different sets of variables, for example, childhood history, adult adjustment, the characteristics of the original offense, and the results of various tests (such as the PCL-R and IQ tests).³³⁸ The resulting variables selected in a majority of subgroups then underwent a final regression analysis.³³⁹ The researchers selected twelve variables for the instrument based on the results of the regression analysis, and each was weighted with reference to the overall base rate in the population, and given a score.³⁴⁰ The

³³³ See *id.* (explaining that multiple regression “involves a variable to be explained—called the dependent variable—and additional explanatory variables that are thought to produce or be associated with changes in the dependent variable”).

³³⁴ See *id.* at 205-06.

³³⁵ See FINKELSTEIN & LEVIN, *supra* note 263, at 350 (giving examples of regression models in law).

³³⁶ See Rubinfeld, *supra* note 324, at 213, 223 (noting that the desirable properties of the least-squares technique include unbiased estimators—so that if the regression were calculated with different samples, the average of the estimates for each coefficient would be the true parameters; consistent estimators—so that if the sample were very large, the estimates would come close to the true parameters; and efficiency, in that the least-squares estimators have the smallest variance among all linear unbiased estimators). “Least-squares minimizes the sum of the squared differences between the actual values of the dependent variable and the values predicted by the regression equation.” *Id.* at 224.

³³⁷ See SPATZ, *supra* note 91, at 101-02 (explaining that when you have many data points that vary in relationship to each other, drawing a line that best represents the data depends on which points are chosen for the equation, and that least-squares is a method to minimize error in drawing a regression line, so that it best fits the data scores when the scores vary together).

³³⁸ See QUINSEY, *supra* note 2, at 146.

³³⁹ See *id.*

³⁴⁰ See *id.* at 147. Each of the twelve variables was assigned a weight of +1 or -1 for every + or - 5% difference from the violence recidivism base rate of 31%. See WEBSTER ET AL., *supra* note 1, at 33 (using the example of the “ever married” factor, where the recidivism rate for subjects who had ever been married was 21%, yielding a 10% difference from the mean recidivism rate of 31%, giving a weight of -2; and comparing this with the “never married” recidivism rate of 38%, which—yielding a 5% increase over the 31% base rate—was given a score of +1). The highest weight was given to the PCL-R score because, of all the variables, it had the strongest correlation with recurring violence. See *id.* Scores ranged from -28 to +33. See QUINSEY, *supra* note 2, at 147. The mean VRAG score for the population tested was close to zero, with a standard deviation of 12.9.

distribution of offenders scores was a normal curve.³⁴¹ This means that roughly as many people fell within the lowest possible risk category (where the probability of recurring violence was close to 0) as fell into the highest risk category (where the probability of recidivism was close to 1.0).³⁴²

E. *Assessing the Link Between Data and Hypothesis: Helpfulness to the Jury (Relevance and Fit)*

Repeated studies of actuarial methods have demonstrated them to be superior to clinical judgment standing alone.³⁴³ Even such instruments however, with their structured reasoning requirements, do not obviate all the problems of human judgment. For example, the risk factor descriptions may be vague, decreasing their reliability.³⁴⁴ Sometimes the factors are not independent, for example anger and the inability to sustain relationships.³⁴⁵ Moreover, the time period that the risk assessment is to cover, the circumstances under which it will be implemented (prison for life in the case of capital sentencing, and the perhaps eventual release into the community after serving a minimum term of the life sentence³⁴⁶), and the individual's motivation to refrain from violence (including motivation to comply with treatment) must all be taken into consideration in assessing risk.³⁴⁷ Yet rarely are they addressed, either in court or in the actuarial instruments. A major problem with each of the risk instruments is its failure to correlate

Id. The standard deviation is the amount that the typical measurement differs from the average. See David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in FEDERAL JUDICIAL CENTER MANUAL ON SCIENTIFIC EVIDENCE (2d ed. 2000) (noting that "Deviations from the average that exceed three or four SDs are extremely unusual"). This high standard deviation remains unexplained. See Kroner & Mills, *supra* note 191, at 486 (noting that this may be due to "inadequacies in conceptualization and, of equal importance, instrumentation").

³⁴¹ See *id.* at 148. A normal curve, or bell-shaped distribution, means that 95% of the distribution lies within two standard deviations of the mean. Rubinfeld, *supra* note 324, at 224. The standard deviation indicates how far two particular data points deviate from the mean. See, e.g., STATISTICS AND THE LAW 409 (Morris H. DeGroot et al. eds., 1986) (explaining the importance of statistics to scientific data, using the example of environmental data).

³⁴² See WEBSTER ET AL., *supra* note 1, at 34.

³⁴³ See Cunningham & Reidy, *supra* note 89, at 28 (citing studies).

³⁴⁴ See Carson, *supra* note 153, at 258 (noting the problem of reliability). For example, even trained clinicians may differ on what exactly is meant by "glibness" (a factor on the PCL-R) or "lack of insight" (a factor on the VRAG and HCR-20).

³⁴⁵ See *id.*

³⁴⁶ If permitted by statute; in some states, life term is without parole.

³⁴⁷ See *id.* at 261.

the risk factors with neuroscience and evolutionary biology.³⁴⁸

Moreover, it is important to bear in mind that risk is a social construct. Although it uses probabilistic analysis and quantification, it is not an exact science.³⁴⁹ Indeed, all science is value-laden, and risk assessment is not different in that regard.³⁵⁰ Using structured analysis however, offers many advantages in human decisionmaking, particularly in light of the difficulty people have in synthesizing differently-weighted likelihoods of varying significance (such as risk factors for violent behavior).³⁵¹ Thus, actuarial instruments may offer a distinct advantage in assessing risk of violent behavior.³⁵²

On the other hand, rare events—and recurring violence is statistically rare, even among violent offenders—are inherently difficult to predict.³⁵³ The most that can be said for any actuarial risk assessment instrument is that it can give a probabilistic estimate of the level of risk for people that share characteristics with the person who is being assessed.³⁵⁴ And the estimate is subject to a great deal of error.³⁵⁵ The major problem for each of the actuarial instruments is that the ultimate question of whether a particular individual ought to be sentenced to death or to life in prison cannot be answered through a risk appraisal.³⁵⁶ Rather it is a political question. The relative costs of imprisoning a person that may ultimately prove to be violent despite a mistaken prediction of nonviolence versus sentencing to death a person who was mistakenly predicted to be violent can only be assessed in

³⁴⁸ See Don Grubin, *Actuarial and Clinical Assessment of Risk in Sex Offenders*, 14 J. INTERPERSONAL VIOLENCE 331, 342 (1999) (explaining that actuarial instruments do not, but ought to, consider “the hardwiring of the brain that could be either genetic in origin or the result of brain injury” as well as experiences “hindering the formation of attachment bonds and the normal development of empathy”). *But see* Harris et al., *supra* note 130, at 197 (attempting to correlate the PCL-R with neuroscience and conceding that much research still needs to be done).

³⁴⁹ See ROYAL SOCIETY, *RISK: ANALYSIS, PERCEPTION AND MANAGEMENT* 7 (1992) (explaining that some subjectivity is always a part of risk assessment).

³⁵⁰ See Beecher-Monas, *supra* note 178, at 1563.

³⁵¹ See J. RICHARD EISER & JOOP VAN DER PLIGT, *ATTITUDES AND DECISIONS* 100 (1988) (observing that human decision “accuracy declines considerably when the number of features or the number of alternatives increases. . . . [and] reliability with which choice rules are used tends to decrease as the decision-maker’s information load increases”).

³⁵² See N. Morris & M. Miller, *Predictions of Dangerousness*, in *CRIME AND JUSTICE: AN ANNUAL REVIEW OF RESEARCH* 1-50 (M. Tonry & M. Miller eds., 1985).

³⁵³ WEBSTER ET AL., *supra* note 1, at 5.

³⁵⁴ See *id.* at 33 (discussing probabilistic estimates of dangerousness).

³⁵⁵ As the VRAG authors explain, “there is a predictive ‘sound barrier’” of some unknown dimension. Although they assert it is greater than 40% accuracy, they do not have the data to substantiate the assertion. QUINSEY, *supra* note 2, at 168.

³⁵⁶ See *id.* at 152-53 (explaining that the question about how to apply the information gleaned from an actuarial instrument is a matter of policy and depends on the relative costs of false positives and false negatives).

terms of values outside the instrument itself.

Paradoxically, although none of the actuarial instruments for predicting risk of recurring violence is highly accurate (especially over the long periods of time that capital juries are concerned about), that is precisely why they should be admissible (and why it could be ineffective assistance of counsel for a defense attorney not to use such expert testimony³⁵⁷). Rather than lumping all convicted defendants into a yes or no dichotomy for future violence, actuarial instruments permit a more sophisticated analysis of the risk.

V. COGNITIVE PSYCHOLOGY: WHY ACTUARIAL INSTRUMENTS MAY ASSIST THE JURY

Future dangerousness testimony is the major means of persuading the sentencing jury that a convicted defendant poses a threat to society and thus merits the death penalty.³⁵⁸ Despite problems with the accuracy of violence predictions, there is some argument that because actuarial predictions provide the best information available, they should be admitted to help guide decisionmakers in their determinations.³⁵⁹ The most hotly debated topic in juror sentencing deliberations—next to the crime itself—is the issue of the defendant’s dangerousness upon return to society.³⁶⁰ Future dangerousness takes precedence in jury deliberations over any mitigating evidence, such as remorse, mental illness, intelligence, or drug/alcohol addiction, and any concern about the defendant’s behavior in prison.³⁶¹ In fact, dangerousness determinations are part of the explicitly authorized grounds for imposing death in a number of states and are required in Texas and Oregon.³⁶² But whether or not future dangerousness is permitted or required by statute, and whether or not it is even

³⁵⁷ See, e.g., *Williams v. Taylor*, 529 U.S. 362 (2000) (failure to present substantial mitigating evidence during the sentencing phase of a capital murder trial was ineffective assistance).

³⁵⁸ See Albertson, *supra* note 273, at 20.

³⁵⁹ The American Bar Association, for example, notes that even though future dangerousness testimony is highly subjective, courts are reluctant to exclude such evidence because it is the “best information available.” PARRY, *supra* note 165, at 49.

³⁶⁰ See John H. Blume et al., *Future Dangerousness in Capital Cases: Always “At Issue”*, 86 CORNELL L. REV. 397, 398 (2001) (observing, on the basis of interviews with over a hundred capital jurors, that “future dangerousness is on the minds of most capital jurors, and is thus “at issue” in virtually all capital trials, no matter what the prosecution says or does not say”).

³⁶¹ See *id.* at 404 (noting results of capital juror interviews).

³⁶² See *supra* notes 12-13 and accompanying text.

mentioned by prosecutors or presented as evidence in the penalty phase of the case, it remains the major focus of the factfinders.³⁶³ This is true regardless of the presence or absence of expert testimony about future dangerousness in the hearing.³⁶⁴

Jurors often believe—incorrectly—that the law requires a death sentence upon a showing of the defendant's future dangerousness.³⁶⁵ Most people do not believe that defendants sentenced to life will actually be required to spend the rest of their lives in prison.³⁶⁶ They fear that a dangerous person will be released into the community; that is a highly motivating factor in choosing between death and a life sentence.³⁶⁷

There are a number of factors about jury deliberations that increase the likelihood that jurors will overestimate the threat of future violence. The most important of these is a lack of objective information about such predictions.³⁶⁸ Base rate errors plague human decisionmaking in general, so there is no reason to suppose capital jurors are any different. Moreover, jurors are seldom informed of the true rate of violent recidivism among murderers released from prison.³⁶⁹ Interviews with capital jurors revealed that

³⁶³ See, e.g., Stephen P. Garvey, *Aggravation and Mitigation in Capital Cases: What Do Jurors Think?*, 98 COLUM. L. REV. 1538, 1560 (1998) (citing studies emphasizing the “pervasive role future dangerousness testimony plays in and on the minds of capital sentencing jurors”).

³⁶⁴ See Blume et al., *supra* note 360, at 404 (reporting that “even in cases in which the prosecution’s evidence and argument at the penalty phase did ‘not at all’ emphasize the defendant’s future dangerousness, jurors who believed the defendant would be released in under twenty years if not sentenced to death were still more likely to cast their final vote for death than were jurors who thought the alternative to death was twenty years or more”). Indeed, it was the explicit recognition of the importance the jury gives to future dangerousness that motivated the Supreme Court to rule that defendants have a constitutional right to be informed of a death penalty alternative if the prosecution alleged future danger as an aggravating circumstance and the alternative is life in prison without parole. See *Simmons v. South Carolina*, 512 U.S. 154 (1994); see also *Kelley v. South Carolina*, 534 U.S. 246 (2002) (reiterating the Court’s earlier holding in *Simmons*); cf. Garvey, *supra* note 363, at 1560 (observing that future dangerousness “appears to be one of the primary determinants of capital-sentencing outcomes”).

³⁶⁵ See James Liginbuhl & Julie Howe, *Discretion in Capital Sentencing Instructions: Guided or Misguided?*, 70 IND. L.J. 1161, 1174 (1995) (43% of surveyed jurors believed the law required a death sentence upon a showing of future dangerousness).

³⁶⁶ See *Simmons*, 512 U.S. at 159 (citing South Carolina survey showing that 92.9% of all jury-eligible adults surveyed believed that a life sentence meant the defendant would be out of prison within 30 years).

³⁶⁷ See, e.g., *id.* at 159 (noting survey in which 75% of those surveyed said that the amount of time actually spent in prison was either “extremely” or “very important” in choosing between life and death).

³⁶⁸ See Sorensen & Pilgrim, *supra* note 29, at 1254 (noting that lack of objective information about future dangerousness predictions is one of the prime reasons jurors overestimate the defendant’s threat of future violence).

³⁶⁹ See *id.* at 1254-55 (noting studies showing that jurors are unaware of the base rates of violent recidivism among murderers).

jurors who sentenced the defendant to death had median estimates of future violence of eighty-five percent, with a fifty percent median estimate that the defendant would commit another murder.³⁷⁰ Yet, empirical studies demonstrate a counter-intuitive decreased base rate for violence among capital commutes in prison.³⁷¹ In one study of 188 death-sentenced prisoners whose sentences were commuted after Furman, for example, over the more than five years after their release, only one killed again, and only six committed violent offenses.³⁷² In studies of capital commutes paroled into the community, twenty percent returned to prison, but only eight to ten percent committed new felonies.³⁷³ Unless jurors are informed about population base rates and how they should use them in making their decision, risk estimates amount to little more than speculation.³⁷⁴

In addition, jurors consistently underestimate the number of years that must be served for a term of "life in prison," which means that the prisoner will be much older and have less opportunity in terms of potential risk period than jurors believe. The decrease in violence and criminal activity with age is a well-established principle of criminology.³⁷⁵ Base rates of violence are far lower after the age of sixty (when most life prisoners would be eligible for parole) than in the twenties.³⁷⁶

Thus, there is ample reason to believe that probabilistic estimates would be helpful to the jury. Violence risk assessment is not a yes/no dichotomy. Rather, it necessitates acknowledging that risk estimates are less than certain and that the base rate of serious violence among capital offenders is quite low. Nor is it enough for experts to use the words "probability" or "risk" in their testimony. In order to be helpful, the expert needs to educate the jury in a scientifically sound manner, and this includes explicitly stating the statistical basis for the opinion.³⁷⁷

Because of the centrality of the dangerousness determination to juror deliberations, it would be helpful to the jury to have information relating to it. It should however, be the best

³⁷⁰ See *id.* at 1269.

³⁷¹ See Cunningham & Reidy, *supra* note 89, at 23 (citing studies).

³⁷² See Sorensen & Pilgrim, *supra* note 29, at 1254-55.

³⁷³ See Cunningham & Reidy, *supra* note 89, at 25 (citing studies and noting that there are now longer minimum sentences that must be served before parole, making the defendants in the studies much younger than current defendants will be when eligible for parole).

³⁷⁴ See *id.* at 23.

³⁷⁵ See *id.* at 31.

³⁷⁶ See *id.* (citing studies).

³⁷⁷ See *id.* at 36-38 (advocating that experts limit their testimony to predominantly statistical analyses in order to avoid going beyond the limits of their scientific expertise).

information available. Currently, actuarial instruments offer the most accurate way of making such predictions, but care must be taken that the expert carefully explain the limits of such testimony.³⁷⁸

It is important that juries be informed that even the most accurate of the actuarial instruments made predictions of dangerousness for people that did not, in fact, later commit acts of violence. Of people who were placed in the “high risk” category under the VRAG, for example, only fifty-five percent actually committed violent acts upon release. That means forty-five percent did not. Had the “high risk” prediction been the basis for a death sentence, nearly half the people sentenced to death would not, in fact, go on to commit any more acts of violence. More importantly, a major failure of each of the actuarial instruments is its failure to take into account what is being learned from neuroscience, evolutionary biology, and enculturation.

CONCLUSION

The very least that ought to be expected of expert evidence adduced at capital sentencing is a demonstration that it is scientifically valid. The rule of law, with its goal of rational truth-seeking, demands no less. *Daubert's* scientific validity requirement is predicated on due process relevance and reliability fairness concerns despite its genesis under the rules of evidence. Expert testimony at capital sentencing must be scientifically valid to meet these concerns. This is a foundational requirement and clinical predictions cannot rise to the mark. Actuarial testimony can barely squeak through. Actuarial predictions unquestionably have shortcomings. Predicting violence, like predicting the weather, is—at best—subject to a large margin of error.³⁷⁹ Nonetheless, however much we gripe about the accuracy of weather forecasts, we frequently find them useful. Some of us—like the fishermen in *The Perfect Storm*³⁸⁰—even make life and death decisions based on these predictions. Most of us would prefer to base our weather-related decisions on meteorologists' statistical analyses rather than the Farmers' Almanac.³⁸¹

³⁷⁸ See, e.g., Monahan & Steadman, *supra* note 239, at 935-36 (weather) (explaining that predictions of the risk of future violence should be modeled explicitly on weather predictions, with all their qualifiers and uncertainties).

³⁷⁹ See Monahan, *supra* note 3, at 914.

³⁸⁰ *Id.*

³⁸¹ To meet *Daubert* standards, an expert must demonstrate scientific validity by a

The more empirical support we can muster for our decisions, the more accurate they are likely to be. Because juries are likely to evaluate the future dangerousness of any criminal defendant, particularly in a capital crime, they should be provided with the most accurate information that can bear on such an assessment. This includes actuarial risk factor studies, analysis of whether those risk factors were present in the defendant's circumstances, and, probabilistically speaking, what that means in terms of the defendant's likelihood of future violence. If the defendant falls within a group in which three percent of the members may commit future violent acts, for example, the jurors are entitled to that information. If the group to which the defendant falls has a fifty-five chance of repeated violent conduct, they are entitled to that information. In conjunction with this information, capital jurors are entitled to hear how opposing experts assess the applicability of the risk factors to the defendant. But if they get any of this information, they should get all of it. That includes the limitations of prediction. To tell only part of the story is wholly misleading and unscientific. Expert testimony with its "imprimatur of science" must fully disclose the limitations of its scientific opinion.

preponderance, a standard the Farmers' Almanac would be unlikely to meet. Meteorologists however, whose predictions are based on repeated statistical studies of weather patterns, ought to be permitted to testify, as long as they disclose their error rates and limitations of their testimony.

