Washington and Lee Law Review

Volume 60 | Issue 2 Article 2

Spring 3-1-2003

The Epistemology of Prediction: Future Dangerousness Testimony and Intellectual Due Process

Erica Beecher-Monas

The Epistemology of Prediction: Future Dangerousness Testimony and Intellectual Due Process

Erica Beecher-Monas*

Table of Contents

---

* 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 go to William G. McLain, Gregory Mitchell, and Theresa Beiner for offering their helpful comments and criticism. Thanks also to Amy Dunn for her able and enthusiastic research assistance.

I. Introduction

Scientific expertise continues to confound the courts years after the Supreme Court's *Daubert*[1] trio was supposed to solve the dilemma of admitting science into legal argument. The resulting confusion is most apparent in capital sentencing proceedings, in which the judiciary has flung wide the gates to wholly unscientific expert testimony. For a democratic system, in which the rule of law attempts to answer the problems of power and freedom by making the law apply to everyone and by providing rational criteria for distinguishing legitimate from illegitimate uses of power, this lack of rationality has consequences. In this Article, I argue that judicial failure to scrutinize expert testimony relating to future dangerousness results in a massive failure of intellectual due process.[2]

---

1. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 592–93 (1993) (instructing the federal judiciary to make admissibility determinations based on an analysis of the scientific validity of the proffered testimony and on whether the testimony "fits" the issues in the case). *Daubert* was followed by *General Electric Co. v. Joiner*, 522 U.S. 136, 144–45 (1997) (reiterating the trial judge's mandate to review testimony for scientific validity and "fit" and allowing a district court to reject expert testimony relying on studies too dissimilar to the facts before it). The Court extended the purview of *Daubert* to technical as well as scientific evidence in the last case in the trilogy, *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999).

2. *See* Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 YALE L.J. 1535, 1539 (1998) (developing the concept of intellectual due process as a norm "that places important *epistemic* constraints on the reasoning process by which legal decisionmakers apply laws to individual litigants . . . [and requires] that the decisionmaking process not be

In a democracy, in which the enunciated goal of the rule of law is the search for truth in a system that aspires to rationality, accurate information is a prerequisite.[3] This goal begs two questions, of course. What is true? And what is rational? I argue with Quine that truth is empirical and that rationality consists of a structured reasoning process relating sensory input to a web of theoretical output.[4] One description of reality is emphatically not just as good as another. That is why courts look to science in the first place: science has

---

arbitrary from an epistemic point of view"). As Professor Wetlaufer explained, "In law, assent is secured through an appeal to reason and logic, through a strong claim to objectivity and certain knowledge, through a voice that claims objectivity and authority." Gerald Wetlaufer, *Rhetoric and Its Denial in Legal Discourse*, 76 VA. L. REV. 1545, 1565 (1990).

3. *See* WILLIAM TWINING, THEORIES OF EVIDENCE: BENTHAM AND WIGMORE 89–90 (1985) (explaining the truth theory of adjudication as the foundation of good government and remarking that "justice absolutely depends upon it"). This kind of statement tends to make people nervous in a postmodern world, where people doubt the achievability of truth, where many believe that truth is contextual, and where different perspectives on truth abound. *See, e.g.*, DENNIS PATTERSON, LAW AND TRUTH 150 (1996) (characterizing postmodernism as emphasizing the idea that "no practice or discourse enjoys a privileged position vis-a-vis others" and asserting that "truth in law is a matter of forms of legal argument"). Some postmodern scholars reject the concept of law as a search for truth entirely, seeing it rather as a contest for power. *See, e.g.*, Carrie Menkel-Meadow, *The Trouble with the Adversary System*, 38 WM. & MARY L. REV. 5, 13 (1996) (arguing that the adversary system is not a search for truth but a contest, the goal of which is to win). However, this Article argues that just because our attempts to discover the truth may be only relatively successful, and just because we may have different perspectives on what truth is, does not mean that we should abandon the search or that the effort to improve the process should be unavailing. For an amusing and enlightening explanation of why both visions (law as a search for truth and law as a contest) may be correct, see Arthur A. Leff, *Law and*, 87 YALE L.J. 989, 1005 (1978) (acknowledging that although the adversary system "does seem more or less well adapted to providing the more or less accurate data needed for the rational operation [of the system and is] largely capable of answering the question 'what happened' at the legally relevant time," there are important ways in which it is also a contest).

4. WILLARD V. QUINE, PURSUIT OF TRUTH 16–19 (1990). This does not mean that fact and value are unrelated or that sensory data determine scientific theory. On the contrary, the data can always support more than one theory. *See generally* WILLARD V. QUINE, *Epistemology Naturalized, in* ONTOLOGICAL RELATIVITY AND OTHER ESSAYS 69 (1969) (explaining that the under-determination of theory by evidence means that the data will always support more than one theory). Science is true, as opposed to metaphysics, religion or astrology, because all questions are ultimately questions of empirical fact, and science answers these questions more successfully. Quine explained: "I see the question of truth as one to be settled within science, there being no higher tribunal." Willard V. Quine, *Comment on Lavener, in* PERSPECTIVES ON QUINE 229 (Robert B. Barrett & Roger F. Gibson eds., 1990). In terms of the reasoning process, Quine does not see that scientific logic is anything special; it is just like common sense, only more careful. *See* WILLARD V. QUINE, *The Scope and Language of Science, in* THE WAYS OF PARADOX AND OTHER ESSAYS 215, 220 (1976) ("The scientist is indistinguishable from the common man in his sense of evidence, except that the scientist is more careful."). For an elegant exposition of Quine's philosophy, see generally Brian Leiter, *Why Quine Is Not a Postmodernist*, 50 SMU L. REV. 1739 (1997).

primacy in describing the world. It works. Reality bites. Descriptive claims need to correspond to the natural world, and the theory articulated for those claims must be articulable, falsifiable, and open to critique.[5] In addition, empirical information about how people actually reach decisions is important in evaluating the truth-seeking and rationality functions of the law.[6]

In our adversary system, the truth-seeking rationality goal of the rule of law forms the basis for evidentiary rules. The basic idea is that the methodologies of the justice system should have truth-generating capacity—a notion of due process.[7] A second consequence of the aspiration to rationality is a concern for accurate evidentiary input: in order to reach a justifiable decision, courts must base reasoning on trustworthy information.[8] A third consequence is that even trustworthy facts must have some logical tendency to prove or disprove an issue in the case.[9] This framework for justice is the inspiration for

---

5. Erica Beecher-Monas, *The Heuristics of Intellectual Due Process: A Primer for Triers of Science*, 75 N.Y.U. L. Rev. 1563, 1584–86 (2000) (describing Karl Popper's theory of science as including the idea that valid theories must be articulable, falsifiable, and open to critique).

6. *See* Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491, 1494 (2001) (arguing that "theorizing is constrained by empirical facts" and that "we cannot craft epistemic norms (norms that would guide our acquisition of knowledge) without empirical information about how the human cognitive apparatus actually works").

7. In other words, "a process reasonably designed to ascertain the truth." JOHN RAWLS, A THEORY OF JUSTICE 239 (1971); *see also* WILLIAM L. TWINING, RETHINKING EVIDENCE 107 (1990) (discussing the rationalist tradition in evidence scholarship and its main epistemological assumption that the purpose of adjudication is to discover an objectively knowable truth, while at the same time acknowledging that "the notion of 'fact' in adjudication is more problematic than the orthodox view suggests . . . . Thus it is misleading to suggest that legal enquiries into questions of fact are value-free"). Even law and economics adheres to the notion of legal process as a search for truth. *E.g.*, Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399 (1978). As Posner put it, judicial error is "a source of social costs and the reduction of error is a goal of the procedural system." *Id.* at 401. That is why, Posner explains, a procedural rule such as the constitutional exclusionary rule "is exceptional, and is recognized—and often bitterly criticized—as such." *Id.*

8. The belief that decisions based on correct information come closer to the truth is the basis of normative epistemology, including "norms governing how individuals should acquire and weigh evidence as well as, ultimately, form beliefs." Allen & Leiter, *supra* note 6, at 1498 (contending that the rules of evidence "structure the epistemic process by which jurors arrive at beliefs about disputed matters of fact at trials").

9. Sometimes this concept is called "materiality," and it is considered to be one of the generative principles of the law of evidence. *See* Robert P. Burns, *Notes on the Future of Evidence Law*, 74 TEMP. L. REV. 69, 70 (2001) (noting that the generative principle of materiality, now subsumed under the relevance requirement, permits into evidence only that evidence that is "of consequence" to the "*legitimate* determination of the action").

the rules of evidence, and a fundamental tenet is that only facts having relevance—rational probative value—should be admissible in the search for truth.[10]

In one important category of proceedings, however, courts toss this framework to the winds. Sentencing hearings have become an evidentiary free-for-all.[11] Particularly in capital sentencing proceedings, in which death is supposed to be different, courts permit juries to hear expert testimony that even the most optimistic could only characterize as not "always wrong."[12] This is a far cry from the truth-generating methodologies supposedly funda-

---

10. A corollary is that all facts that have rational probative value should be admissible unless forbidden under a competing concern of the justice system (for example, the improper uses of state power implicated in the exclusionary rule). TWINING, *supra* note 7, at 152. *But see* Randolph N. Jonakait, *The Origins of the Confrontation Clause: An Alternative History*, 27 RUTGERS L.J. 77, 82 (1995) (contending that evidence law grew out of the adversarial system as part of the "Sixth Amendment rights that affirmatively grant an accused the opportunity for meaningful defense advocacy"). The doctrines of relevance and probativity are expressed as follows under the Federal Rules of Evidence:

> "Relevant evidence" means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

FED. R. EVID. 401. And:

> Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

FED. R. EVID. 403.

11. Among the various types of evidence that are excluded from trials under rules of evidence but constitutionally permitted at sentencing are victim impact statements, acquitted conduct, and unadjudicated conduct. *See, e.g.*, United States v. Watts, 519 U.S. 148, 154–55 (1997) (per curiam) (stating that a "not guilty" verdict does not preclude admissibility of conduct underlying the acquitted charge during sentencing, as long as the prosecutor proves the conduct by a preponderance of the evidence); Romano v. Oklahoma, 512 U.S. 1, 10 (1994) (finding a prior death sentence admissible in capital sentencing); Payne v. Tennessee, 501 U.S. 808, 827 (1991) (declaring victim impact evidence constitutionally admissible under the Eighth Amendment). Hearsay is the most prevalent category of evidence that courts both proscribe during the guilt phase of the trial and widely admit at sentencing. *See, e.g.*, Williams v. New York, 337 U.S. 241, 244–52 (1949) (finding it constitutionally permissible to rely on a presentence report because of the need for a broad spectrum of information at sentencing). For an excellent discussion of the applicability of the rules of evidence at sentencing, see Deborah Young, *Fact-Finding at Federal Sentencing: Why the Guidelines Should Meet the Rules*, 79 CORNELL L. REV. 299, 302–03 (1994) (arguing that the Federal Rules of Evidence should apply at federal sentencing).

12. Barefoot v. Estelle, 463 U.S. 880, 900 (1983) (refusing to exclude future dangerousness testimony because the defense could not show that "psychiatrists are always wrong with respect to future dangerousness, only most of the time").

mental to due process and at the opposite end of the spectrum from what currently happens in civil trials, in which experts must demonstrate the reliability of their testimony.[13]

Outside of sentencing proceedings, courts have approached the issue of reliability with respect to scientific testimony in two basic ways. The first is the *Frye*[14] general consensus approach, in which courts deem scientific testimony trustworthy if a majority of scientists agree that it is—assigning the gatekeeping task to scientists.[15] Alternatively, the Federal Rules of Evidence now assign the gatekeeping responsibility to judges.[16] My position, as I have

---

13. *See generally* Julie G. Shoop, *Judges Are Gaining Confidence in Assessing Expert Evidence, Study Finds*, 38 TRIAL 92 (2002) (discussing a report by the Rand Institute for Civil Justice, which found that *Daubert* has had a significant impact on the admissibility of expert testimony in civil trials, so that judges are closely scrutinizing relevance and reliability, resulting in a dramatic increase in the percentage of excluded expert testimony in products liability trials and a surge in summary judgments against the plaintiffs). This increased scrutiny of scientific evidence has not affected criminal trials nearly so dramatically. *See, e.g.*, Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 78-82 (1998) (describing the "maelstrom" of unresolved issues created by or unaddressed in *Daubert* and the changes that it entailed); Paul C. Gianelli, *Scientific Evidence in Civil and Criminal Cases*, 33 ARIZ. ST. L.J. 103, 113–17 (2001) (discussing the myriad instances in which criminal judges fail to take *Daubert* seriously).

14. Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

15. *See id.* at 1014 (explaining that scientific testimony must "be sufficiently established to have gained general acceptance in the particular field in which it belongs"). For years, courts based the validity determination on scientific consensus, under which scientists themselves would decide—more or less by majority rule—whether a given assertion was scientifically valid. Only if the scientific community generally accepted a given scientific technique or conclusion would it be admissible in court. This general acceptance rule of *Frye* was a bad idea because general acceptance alone simply is not a measure of validity; the scientific community has rubber-stamped lots of inane ideas and denigrated lots of good ones for reasons that had absolutely nothing to do with "scientific validity" and much to do with politics, context, and culture. The kind of research done in the first place depends on the availability of funding, which may be politicized and is, in any event, scarce. Once the hurdles of funding have been overcome, publication is by no means assured. Far more papers are written than ever will be published, often consigning research results to obscurity. Moreover, interpretation is a human activity, not a process of nature. The collection and explanation of scientific data is a rhetorical act. In essence, the effect of leaving the validity determination to scientific consensus is to cede the power to determine political questions of culpability to an extra-judicial (and extra-governmental) power.

16. Rule 702 now provides:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of

argued elsewhere, is that judges are the proper gatekeepers for a number of reasons, but primarily because the structured inquiry process of *Daubert* gives judges a framework for analysis and a requirement to justify—rationalize—their decisions.[17] Whichever approach they use, however, trustworthiness, i.e., relevance, is a crucial aspect of admissibility. And courts use neither the *Frye* approach nor the *Daubert* approach (codified in the federal rules), nor any other trustworthiness inquiry, in one critical area of sentencing proceedings: expert testimony about the defendant's future dangerousness.

This result is due in large part to the Supreme Court's decision in *Barefoot v. Estelle*.[18] This Article argues that the Court decided *Barefoot* wrongly, both as a matter of evidentiary due process and because it was empirically wrong about the ability of the adversary system to sort out the reliable from the unreliable expert testimony.[19] Moreover, this Article argues that *Barefoot* is egregiously wrong-headed by current standards for relevance[20] and that *Barefoot*'s effect on capital sentencing proceedings has been pernicious and pervasive, undermining basic rule-of-law precepts. Perhaps it is true, as Justices Blackmun and Powell argued, that courts cannot administer the death penalty in a way that meets constitutional requirements.[21]

---

the case.

FED. R. EVID. 702. In addition to federal courts, in which the federal rules are mandatory, thirty-eight states have adopted the rules as the basis for their state procedure. Daniel A. Krauss & Bruce D. Sales, *The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing*, 7 PSYCHOL. PUB. POL'Y & L. 267, 270 (2001).

17. *See generally* Beecher-Monas, *supra* note 5 (outlining a heuristic to guide judges in deciding admissibility questions).

18. *See* Barefoot v. Estelle, 463 U.S. 880, 892 (1983) (permitting experts to testify about future dangerousness as a constitutional matter).

19. Krauss & Sales, *supra* note 16, discusses research suggesting that jurors are incapable of differentiating more scientifically valid expert testimony from less accurate testimony.

20. Notably, no one persuasively argues that the testimony at issue in *Barefoot* could meet *Daubert* standards. *See, e.g.*, 1 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 2 (2d ed. 2002) (arguing that *Barefoot* and *Daubert* are irreconcilable).

21. *See* Callins v. Collins, 510 U.S. 1141, 1145-46 (1994) (Blackmun, J., dissenting) (observing that "the inevitability of factual, legal, and moral error gives us a system that we know must wrongly kill some defendants, a system that fails to deliver the fair, consistent, and reliable sentences of death required by the Constitution"); Jeffrey L. Kirchmeier, *Aggravating and Mitigating Factors: The Paradox of Today's Arbitrary and Mandatory Capital Punishment Scheme*, 6 WM. & MARY BILL RTS. J. 345, 347 (1998) (noting that "Justice Powell came to a similar conclusion after his retirement"). Notably, two states, Illinois and Maryland, have reached similar conclusions and have placed a moratorium on death penalty prosecutions because of due process concerns. *See* Dirk Johnson, *Illinois, Citing Verdict Errors, Bars*

If, as this Article contends, the test by which courts should judge an evidentiary practice is whether it increases the likelihood that the truth—defined as correspondence to the real world—will be attained, expert future dangerousness testimony fails to make the grade. This Article argues that although valid reasons may justify abandoning strict adherence to the rules of evidence when it comes to sentencing proceedings—and I am skeptical about this proposition—at a very minimum, expert testimony should be admissible only if it is scientifically valid.[22]

The rule that this Article proposes is a constitutionalization of *Daubert*, at least with respect to death penalty proceedings. The judge's task is not to decide whether the expert's proposition is correct, but to decide whether, given the current state of knowledge about the issue, as propounded by competing experts, it is sufficiently trustworthy to come into evidence in order to prove or disprove a legal issue (in this case, the risk of future violence).[23] In sum, the reason that we need gatekeepers is to ensure that the statements offered into evidence comport with permissible legal theories, embedded as they are in cultural systems of belief, assumptions and claims about the world.[24] Although what we seek to know are the facts, as Quine explains, facts are inevitably theory-laden.[25] Therefore, in an adversary system, the judge's role is to manage coherence by reference to what is relevant to the legal determination. And *Daubert* is unequivocal that relevance in the context of expert testimony means scientific validity.

Executions, N.Y. TIMES, Feb. 1, 2000, at A1 (describing the decision to bar executions in Illinois).

22. Notably, one-third of the prisoners exonerated by the Actual Innocence Project had been convicted on the basis of "tainted or fraudulent science." BARRY SCHECK ET AL., ACTUAL INNOCENCE 246 (2000).

23. Contrary to Professor Brewer's metaphor, the judge's task is not like having to decide which of "two groups of expert mathematicians [who] disagree about a complex mathematical question—say, whether Princeton mathematician Andrew Wiles really did solve 'Fermat's Last Theorem,' which no mathematician had been able to prove since Louis Fermat first propounded it about 360 years ago"—is correct. Brewer, *supra* note 2, at 1538. The judge's task is not to decide whether Wiles's solution is correct, but to decide whether, given the current state of knowledge about the courtroom analog to Wiles's solution, as propounded by competing experts, it is sufficiently trustworthy to come into evidence in order to prove or disprove a legal issue (in this case, risk of future violence).

24. For the theory that knowledge is embedded in a network of meanings, see WILLARD V. QUINE, *Two Dogmas of Empiricism*, *in* FROM A LOGICAL POINT OF VIEW 20–46 (2d ed. 1980).

25. WILLARD V. QUINE, *Epistemology Naturalized*, *in* ONTOLOGICAL RELATIVITY AND OTHER ESSAYS 83 (1969), explains the "reciprocal containment" of "epistemology in natural science and science in epistemology." Thus, Quine is committed to empiricism, but finds that knowledge and theory are inseparable.

This Article argues that relevance is not "merely" a matter of evidentiary rules, but rather a constitutional minimum, a requirement of due process and a fundamental fairness requirement of the rule of law. Contrary to the Court's contention in *Barefoot*, the adversary process cannot be trusted "to sort out the reliable from the unreliable evidence and opinion about future dangerousness."[26] Rather, as the Supreme Court explained in *Daubert*, the requirement that expert testimony be helpful to the jury, "supported by appropriate validation—i.e., 'good grounds,' based upon what is known," is a condition of relevance.[27] This Article examines the dynamics of jury decisionmaking in the context of capital sentencing decisions to explain why the relevance requirement is a constitutional concern of fundamental fairness and how that requirement plays out with respect to expert testimony on future dangerousness.

This Article proceeds in six parts. Following this Introduction, Part II discusses the prevalence of future dangerousness evidence in capital sentencing hearings. Part III addresses the Supreme Court's decision in *Barefoot v. Estelle*[28] and the tension between *Barefoot* and the Supreme Court's *Daubert* trilogy of cases (and their codification in the Federal Rules of Evidence).[29] Part IV addresses what is wrong with clinical predictions and why admitting them into capital sentencing proceedings violates basic premises of the rule of law and the adversary system. In terms of promoting the acquisition of knowledge for capital jurors to make their life-or-death decision, permitting experts to expound bunkum does not comport with fundamental aspects of the rule of law. Part IV thus advocates applying the relevance and reliability standards of *Daubert* to expert testimony on future dangerousness.

Part V draws on data from jury decision studies, in particular the Capital Juror Project, and from insights into cognitive biases and human decisionmaking to analyze the relevance, in terms of scientific validity, of actuarial instruments currently used to assess the risk of future violence and concludes that actuarial instruments, although they have shortcomings, offer an improvement over the unaided judgment of juries and over clinical predic-

---

26. Barefoot v. Estelle, 463 U.S. 880, 901 (1983).

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

28. *Barefoot*, 463 U.S. at 892 (permitting experts to testify about future dangerousness as a constitutional matter).

29. *See* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 141 (1999) (extending the scope of the *Daubert* inquiry to technical as well as scientific evidence); Gen. Elec. Co. v. Joiner, 522 U.S. 136, 144–45 (1997) (reiterating the trial judge's mandate to review testimony for scientific validity and "fit"); *Daubert*, 509 U.S. at 592–93 (instructing the federal judiciary to make admissibility determinations based on an analysis of the scientific validity of the proffered testimony and on whether the testimony "fits" the issues in the case).

tions of violence. This Article concludes that judicial gatekeeping standards for scientific evidence, as outlined by the *Daubert* trio of cases, are an essential component of due process and that the trustworthiness of expert scientific testimony—in a system that aspires to rationality—is a minimum prerequisite.

II. Expert Future Dangerousness Testimony in Capital Sentencing Proceedings

Expert testimony about future dangerousness currently takes two forms: clinical predictions and actuarial predictions. By far the most common form of testimony is the clinical prediction. In this testimony, the expert, usually a psychiatrist but sometimes a psychologist, proffers an opinion based on a courtroom hypothetical, with or without a prior examination of the defendant. Even when clinicians have the opportunity for an extensive examination, however, studies show that clinical predictions are highly inaccurate. A number of factors explain this inaccuracy. Clinical decisionmakers tend to assume erroneously the representativeness of events by ignoring sample sizes[30] and base rates. Ignoring base rates is a particular problem in predicting violence when the base rate of violent behavior is low overall and varies among different population subgroups.[31]

In addition, these studies show both that clinicians tend to think that they have more information than they really do[32] and that they are poor at making extreme judgments.[33] Clinical judgments tend to ignore the well-known difficulty in predicting statistically rare events (like violence).[34] Stereotypes and prejudices are just as likely to taint the decisions of clinicians as those of lay people.[35] As a result, clinicians are no better than lay people in making

---

30. Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 COGNITIVE PSYCHOL. 430, 433 (1972). Decisionmakers do not consciously employ these shortcuts, but rather they operate on a subliminal level to affect decision. *Id.*

31. *See* VERNON L. QUINSEY ET AL., VIOLENT OFFENDERS: APPRAISING AND MANAGING RISK 60 (1998) (explaining how the statistical base-rate problem affects predictions of dangerousness).

32. *Id.* at 56 ("[P]eople often think they have more information than they actually have, they are therefore more willing to make more extreme judgments than warranted.").

33. *See* Hillel J. Einhorn & Robin M. Hogarth, *Confidence in Judgment: Persistence of the Illusion of Validity*, 85 PSYCHOL. REV. 395, 396 (1978) (noting that clinicians tend to have more confidence in predictive variables with extreme values than is warranted).

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

35. *See* CHRISTOPHER WEBSTER ET AL., THE VIOLENCE PREDICTION SCHEME: ASSESSING DANGEROUSNESS IN HIGH RISK MEN 28 n.5 (1994) (quoting Judge Bazelon's opinion that a

these predictions.[36] Moreover, no information is available about the individual error rate of the particular expert proffering a future dangerousness opinion.[37] Although some experts are undoubtedly better at diagnosis and prediction than others, 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 thus is virtually untestable.[38]

Actuarial instruments attempt to correct these deficiencies by relying on statistically analyzed data rather than personal experience.[39] These instruments attempt to counter human cognitive error by taking into account the interrelationship of various risk factors and the population base rates, and by assigning weights to the individual risk factors. Repeated studies of actuarial methods have demonstrated them to be superior to clinical judgment standing alone.[40] Actuarial instruments do not abandon human judgment; they simply structure it into a formal reasoning process.[41] Using structured analysis offers

---

psychiatrist's internal values could undermine his testimony).

36. *See* QUINSEY ET AL., *supra* note 31, at 62 (reporting in a study assessing predictions of violence that "lay persons and the clinicians had few differences of opinion"); 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) (noting that "expert judgments that are clinically derived, as opposed to actuarially derived, are as susceptible to error as lay judgements").

37. *See* WEBSTER ET AL., *supra* note 35, 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).

38. *See* Shuman & Sales, *supra* note 36, at 1227 (noting the problem of unvalidated theories and skills).

39. Although some evidence suggests that a multidisciplinary team may be able to rival the accuracy of actuarial instruments, courts are unlikely to employ such teams in capital sentencing determinations. *See* J. Fuller & J. Cowan, *Risk Assessment in a Multidisciplinary Forensic Setting: Clinical Judgement Revisited*, 10 J. FORENSIC PSYCH. 276, 286 (1999) (acknowledging that a multidisciplinary team may provide increased accuracy approaching that of actuarial instruments). For a description and discussion of actuarial instruments used to predict future dangerousness, see generally Erica Beecher-Monas & Edgar Garcia-Rill, *Chaos at the Edge of Danger: Predicting Violent Behavior in a Post-*Daubert *World*, 24 CARDOZO L. REV. ___ (forthcoming 2003).

40. *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, 28 (1999) (citing studies); Mark Dolan & Mary Doyle, *Violence Risk Prediction: Clinical and Actuarial Measures and the Role of the Psychopathy Checklist*, 177 BRIT. J. PSYCH. 303, 303 (2000) (listing assessment instruments used for violence risk assessment in the mentally disordered and citing studies). Dolan and Doyle note that "structured/systematic approaches to violence risk prediction provide a more accurate and transparent record of the risk factors considered and the rationale behind decisions taken." *Id.* at 309.

41. *See* QUINSEY ET AL., *supra* note 31, at 65 ("[H]uman judgments applied in a very

many advantages in human decisionmaking, particularly in light of the difficulty that people have in synthesizing differently weighted likelihoods of varying significance, such as risk factors for violent behavior.[42]

These instruments are only beginning to find their way into capital sentencing proceedings, although prisons and other forensic settings, such as hospitals, widely use them. In *United States v. Barnette*,[43] the prosecution expert used an actuarial instrument, the Psychopathy Checklist Revised (PCL-R), to substantiate his opinion that the defendant posed a future danger.[44] The court upheld the admissibility of this testimony under *Daubert* (although it declined to decide whether a *Daubert* analysis was required).[45] The court found that the prosecution expert's testimony met *Daubert*'s reliability standard because he had based his opinion on "observations of Barnette's behavior; the actuarial approach; and the research on predicting future dangerousness."[46] The court did not, however, analyze the scientific validity of the actuarial approach. Thus, although the court invoked *Daubert*, it circumvented *Daubert*'s analysis requirements.

III. Evidentiary Contradictions: Barefoot *and* Daubert

A. Barefoot *and the Regime of Evidentiary Federalism*

One can trace the failure of courts to address issues of scientific validity at sentencing directly to *Barefoot v. Estelle*. In *Barefoot*, despite its previously

---

structured way play a large role in the actuarial prediction of violence."); Dolan & Doyle, *supra* note 40, at 304 (observing that "[s]tructured clinical judgment represents a composite of empirical knowledge and clinical/professional expertise").

42. *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"); Fuller & Cowan, *supra* note 39, at 286 (finding that empirical data demonstrates that such structured analysis improves decisionmaking considerably). Even these instruments do not obviate all the problems of human judgment. For example, the risk factor descriptions may be vague, decreasing their reliability. *See* David Carson, *A Risk Management Approach to Legal Decision-Making About 'Dangerous' People*, *in* LAW AND UNCERTAINTY: RISKS AND LEGAL PROCESS 255, 258 (Robert Baldwin ed., 1997) (noting the problem of reliability and overlap in risk factors). For example, even trained clinicians may differ on what exactly is meant by "glibness" (a factor on the PCL-R, an actuarial instrument) or "lack of insight" (a factor on the VRAG and the HCR-20, two other actuarial instruments). Sometimes the factors are not independent, for example, anger and the inability to sustain relationships. *Id.*

43. United States v. Barnette, 211 F.3d 803 (4th Cir. 2000).

44. *See id.* at 815 (discussing *Daubert*, then stating that the decision to admit evidence of PCL-R was not a clear error of judgment sufficient to reverse the lower court).

45. *Id* at 815–16.

46. *Id.* at 816.

articulated concern for providing the jury with accurate information with which to make its decision, the Supreme Court upheld expert testimony about future dangerousness for which even the majority could not find any scientific support.[47] 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.[48]

The Court based its reasoning not only on precedent, but also on the rules of evidence, "generally extant at the federal and state levels," which, according to the Court, "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."[49] The Court thus reasoned that the testimony was admissible because the rules of evidence deemed it admissible, and emphasized the ability of the adversary system to weed out inaccurate information.[50] At stake were federalism notions that the Court would not interfere unless the state practice offended constitutional minimums. The rules of evidence at the time did not require expert testimony to undergo any scrutiny for scientific validity.[51]

---

47. Barefoot v. Estelle, 463 U.S. 880, 899 (1983). After acknowledging the opposition to future dangerousness testimony set out in the amicus brief of the American Psychiatric Association stating that the unreliability of clinical predictions of dangerousness was an "established fact," the only support that the majority could find for such testimony was one study showing "some predictive validity" and a statement by a researcher (relied on by the state experts). Although "the 'best' clinical research currently in existence indicates that *psychiatrists and psychologists are accurate in no more than one out of three predictions of violent behavior*," the researcher stated that "there *may be* circumstances in which prediction is both empirically possible and ethically appropriate" and hopes that "future research would clarify the issues." *Id.* at 899 n.7 (emphasis added). The presence of widespread "professional doubts about the usefulness of psychiatric predictions" should only make it that much easier to mount a challenge to the experts. *Id.* Because the majority was "not persuaded that such testimony is almost entirely unreliable," it would be up to the jury to "separate the wheat from the chaff." *Id.* Contrast this analysis with the Court's analysis in *General Electric Co. v. Joiner*, a civil case construing the Federal Rules of Evidence, in which the Court insisted that a trial court must examine the expert's methodology and techniques for consistency with the expert's conclusions and with the facts of the case at hand. 522 U.S. 136, 145–46 (1997) (declining to admit expert testimony based on sophisticated animal and epidemiology studies because the studies lacked "fit").

48. *Barefoot*, 463 U.S. at 901 ("Neither the petitioner nor the Association suggests that psychiatrists are always wrong with respect to future dangerousness, only most of the time.").

49. *Id.* at 898.

50. *See id.* (describing checks in the adversary system such as cross-examination and contrary evidence).

51. *See* Michael H. Gottesman, *From* Barefoot *to* Daubert *to* Joiner: *Triple Play or Double Error?*, 40 ARIZ. L. REV. 753, 755 (1998) (noting that expert testimony at the time was admissible as long as the expert was qualified); *see also* 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

Two psychiatrists testified in *Barefoot* that the defendant "would probably commit further acts of violence and represent a continuing threat to society."[52] 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, but had no history of violent crime.[53] Instead, the experts based their testimony on a hypothetical question based on the crime and the defendant's conduct.[54] The Supreme Court upheld the admissibility of such testimony, remarking that disallowing it would be like "disinvent[ing] the wheel."[55] Because courts had traditionally admitted such testimony, the Court refused to overturn its precedent.[56] The Court acknowledged the opposition of the American Psychiatric Association (the Association) to future dangerousness testimony because of its extreme unreliability.[57] Although the Association explained that no one, including psychiatrists, can predict with any degree of reliability that an individual will commit other crimes in the future,[58] the Court noted that the Association did not claim that psychiatrists were *always* wrong with respect to future dangerousness predictions—only that they were

---

admissibility for expert testimony—reliability—is also the paramount concern in addressing the constitutionality of capital sentencing procedures").

52. *Barefoot*, 463 U.S. at 884.

53. *Id.* at 917 (Blackmun, J., dissenting).

54. *See id.* at 918–19 (Blackmun, J., dissenting) (explaining the psychiatrist's method). 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 in which the defendant had lived over a period of ten years, unemployment during the two months preceding the crime, drug use, boasting to acquaintances of plans to commit crimes, shooting a police officer without provocation from a distance of six inches, and acting after the crime as though nothing unusual occurred. Brief of the American Psychiatric Association as Amicus Curiae at 5, Barefoot v. Estelle, 463 U.S. 880 (1983).

55. Barefoot v. Estelle, 463 U.S. 880, 899 (1983).

56. *See id.* (relying on precedent stating that a defendant's likelihood to commit further crimes is a constitutional criterion for imposing the death penalty). The Court expounded that "if it is not impossible for even a lay person sensibly to arrive at that conclusion, it makes little sense, if any, to submit that psychiatrists, out of the entire universe of persons who might have an opinion on the issue, would know so little about the subject that they should not be permitted to testify." *Id.* at 896–97. The idea is that courts permit experts to testify because they can assist lay persons making a decision; otherwise they would be invading the province of the jury. But ample evidence suggests—and the Court was aware of it through the amicus brief—that the experts are no better than lay jurors in making this decision. *See, e.g.*, QUINSEY ET AL., *supra* note 31, at 62 (observing that clinicians and lay people had little difference of opinion in making violence predictions).

57. *Barefoot*, 463 U.S. at 899.

58. *Id.* at 896–99.

wrong more often than not—and therefore it would not exclude such testimony because it comported with the state's rules of evidence.[59]

State rules of evidence have changed since *Barefoot* thanks to the Supreme Court's ruling in *Daubert*. Even states that elected not to follow *Daubert* and retained a general consensus standard have modified their approach to expert testimony. What has not changed and what the Supreme Court did not mention is that the rules of evidence do not generally apply to sentencing hearings. Thus, if *Daubert* requires rethinking expert testimony at capital sentencing, the reason must be that *Daubert*'s foundation is a constitutional one. This Article argues that *Daubert*'s foundation is the relevance and reliability of expert testimony and that this foundation is constitutionally based.

B. Daubert, *Its Progeny, and the Federal Rules of Evidence in Death Penalty Proceedings*

Justice Blackmun, who wrote the dissenting opinion in *Barefoot*, wrote the majority opinion in *Daubert* and with it transformed the jurisprudence of expert testimony. Explicitly addressing the Federal Rules of Evidence, *Daubert* requires judges to examine the empirical basis of statements made by experts in federal courts. *Daubert* held that the scientific validity and the "fit" of expert testimony to the facts in the case are questions of reliability and relevance.[60] In two subsequent cases, *General Electric Co. v. Joiner*[61] and *Kumho Tire Co. v. Carmichael*,[62] the Court reiterated the *Daubert* standards, expounded on its notion of "fit," and explained that judges not only must evaluate the scientific validity of testimony based on the traditional "hard" sciences, but also must evaluate the validity of expert testimony based on the "soft" sciences, such as psychology.[63] Thus, *Daubert*'s general principles

---

59. *Id.* at 904.

60. *See* Paul C. Giannelli, *Daubert: Interpreting the Federal Rules of Evidence*, 15 CARDOZO L. REV. 1999, 2021 (1994) (observing that "*Daubert* required a higher standard of admissibility for money damages than *Barefoot* required for the death penalty"). Although commentators have argued that *Daubert* is technically consistent 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").

61. Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).

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

63. *See id.* at 149 (finding that *Daubert* applies to all expert testimony described in

apply to all expert testimony.[64] Congress subsequently amended the Federal Rules of Evidence to codify these cases.[65]

These changes have profoundly affected not only the federal courts but also state courts.[66] Even jurisdictions that eschew the *Daubert* standard in favor of the general acceptance standard of *Frye* are beginning to insist that expert testimony meet standards of scientific validity.[67] Thus, even when courts do not follow the Supreme Court's standard—that is, in many state courts—they have an increased awareness that whatever evidence they consider should be based on a sound empirical foundation.

The rationale for a gatekeeping requirement is based on relevance.[68] Unreliable testimony is not relevant because it cannot assist the jury, and assisting the jury is the only reason for admitting expert testimony. Four "general observations" guide the inquiry into scientific validity.[69] They are testability; peer review and publication; the existence of methodological standards, including the error rate of the methodology; and general acceptance.[70] The overall goal of these "flexible" guidelines is to evaluate expert testimony by the standards that scientists themselves use to critique each other's work. Relevance is thus the capstone of *Daubert*'s scientific validity inquiry. The relevancy requirement functions as the primary control of the court over what information parties may present. As Professor Damaska explains, "Anglo-American criteria of relevancy make the factual basis of a decision closer to social reality, where fact and value are intertwined."[71] In an adversarial system, truth is more a matter of perspective than in the inquisitorial system, and the court controls the flow of information presented by the parties as a way of controlling social conflict.[72]

Federal Rule of Evidence 702). Engineering testimony was at issue in *Kumho Tire*. *Id.* at 137.

64. *See id.* at 148 (finding that courts do not limit the *Daubert* rationale to "scientific" knowledge).

65. *See* FED. R. EVID. 702 (incorporating *Daubert*'s language).

66. *See, e.g.*, United States v. Norwood, 939 F. Supp. 1132, 1134 (D.N.J. 1996) (applying the *Daubert* test).

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

68. *See* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 147 (1999) (finding that the gatekeeping function applies to all experts because *Daubert*'s focus was on the relevance of the testimony, not its scientific nature).

69. *See* Daubert v. Merrell Dow Pharms., 509 U.S. 579, 594 (1993) (characterizing the inquiry as a "flexible one").

70. *Id.* at 593–94.

71. Mirjan Damaska, *Presentation of Evidence and Factfinding Precision*, 123 U. PA. L. REV. 1083, 1104 (1975).

72. *See id.* at 1105 (contrasting the adversary and inquisitorial systems of adjudication).

The profession has overwhelmingly castigated future dangerousness testimony based on clinical judgment alone (and so it fails the peer review, publication, and general acceptance prongs of *Daubert*). Because such predictions are wrong more often than right, they cannot meet the error rate inquiry.[73] 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 finding future dangerousness testimony admissible was that even a lay person could testify to future dangerousness.[74] Unlike the lay persons involved, however, these experts had no personal knowledge of the defendant.[75] The *Barefoot* experts based their testimony entirely on the defendant's conduct at trial and the facts of the crimes, as most such experts still do in Texas. This basis was enough, according to one of the experts, to demonstrate future dangerousness with "one hundred percent" accuracy.[76] That statement alone was enough to make his testimony unreliable.[77]

One of the reasons for permitting all relevant evidence into a sentencing hearing, regardless of whether it can meet the rules of evidence (provided that it is sufficiently reliable), is to assist the jury in making an individualized determination of whether the death penalty is appropriate under 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 actually assist the factfinder.[78] The Court pointed out in *Joiner* that conclu-

---

73. Although they made bald assertions that they were invariably accurate, the experts in *Barefoot* offered no substantiation for their claims. *See* Barefoot v. Estelle, 463 U.S. 880, 918–19 (1983) (Blackmun, J., dissenting) (recounting the doctor's claims and the lack of substantiation from the trial).

74. *See id.* at 896–97 (relying on precedent that allowed lay persons to judge future dangerousness).

75. *See id.* at 917 (Blackmun, J., dissenting) (noting that neither doctor ever personally examined Barefoot).

76. *Id.* at 919 (Blackmun, J., dissenting) (emphasis deleted).

77. No one can testify with "one hundred percent" accuracy. In addition, this expert had no empirical data to support his highly improbable statement. Such hyperbole by experts is by no means uncommon. A Texas psychiatrist, who by 1992 had participated in 144 capital cases, testified in each of them that he was, with medical and scientific certainty, sure that the defendant would kill again. *See* Joseph T. McCann, *Standards for Expert Testimony in New York Death Penalty Cases*, 68 N.Y. ST. BAR J. 30, 30 (1996) (outlining the prevalence of improper assessments of future dangerousness). One of those condemned was later found to be innocent of the crime. *Id.* at 32.

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

sions and methodology must have a valid connection.[79] 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.[80] The overwhelming majority of courts that have addressed the issue since *Barefoot* have simply found its constitutionality beyond question. For example, the Texas Court of Criminal Appeals in 1998, addressing the scientific validity of future dangerousness testimony, found it to be sufficiently reliable.[81]

But a few muted stirrings of unease have occurred. For example, the U.S. Court of Appeals for the Fifth Circuit acknowledged the issue of whether *Daubert* implicitly overruled *Barefoot*, although it declined to reach that issue in *Tigner v. Cockrell*.[82] In addition, concurring in the Fifth Circuit's per curiam opinion in *Flores v. Johnson*,[83] Judge Garza excoriated the Texas courts' use of expert future dangerousness testimony (but found himself bound nonetheless by *Barefoot*).[84]

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 testified that a defendant did not pose a future danger.[85] 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

---

79. *See* Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146-47 (1997) (discussing the link between conclusions and methodology).

80. *See* WEBSTER ET AL., *supra* note 35, at 17–21 (observing that the political and legal pressures on experts to predict violence in the United States and Canada are overwhelming).

81. *See* Nenno v. State, 970 S.W.2d 549, 560 (Tex. Crim. App. 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. 1999). "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." *Id.* at 561.

82. *See* Tigner v. Cockrell, 264 F.3d 521, 526–27 (5th Cir. 2001) (dismissing the case because a decision on collateral review would have violated the non-retroactivity principle).

83. Flores v. Johnson, 210 F.3d 456 (5th Cir. 2000) (per curiam).

84. *See id.* at 464 (Garza, J., concurring) (castigating the "inadequacy of the science" behind expert future dangerousness testimony).

85. *Id.* at 462 (Garza, J., concurring).

right."[86] Moreover, when considered in juxtaposition with the strict admissibility requirements for most expert testimony—especially in civil trials, in which the stakes are much lower—Dr. Griffith's testimony became strikingly inadequate.[87]

The *Barefoot* decision permitted experts to testify about 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."[88] Indeed, faced with such testimony, "juries are almost always persuaded."[89] 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 into evidence.[90] Nonetheless, the court upheld Flores's death sentence.

The reliability standard articulated in *Daubert* and its progeny should apply to future dangerousness evidence. Juries give great credence to expert testimony, and the scientific literature evaluating the predictive value of clinical judgments about future violence has shown that these expert predictions are no better than lay judgments. If courts applied *Daubert* standards to the kinds of clinical predictions currently offered in our courts, they would not admit the predictions 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, in which the jury hearing the evidence might very well impose the death penalty. It seems extraordinarily anomalous that courts should require more to demonstrate relevance and reliability in civil cases than in criminal cases.

---

86. *Id.* (Garza, J., concurring) (quoting Barefoot v. Estelle, 463 U.S. 880, 920 (1983) (Blackmun, J., dissenting)).

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

88. *Id.* at 466 (Garza, J., concurring). Judge Garza opined that admitting a psychiatric prediction of dangerousness was akin to permitting a phrenologist—the example that 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 (Garza, J., concurring) (citing Gen. Elec. Co. v. Joiner, 522 U.S. 136, 153 n.6 (1997) (Stevens, J., concurring)). Judge Garza noted that "the phrenologists' testimony appears no less scientific." *Id.* (Garza, J., concurring).

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

90. Flores v. Johnson, 210 F.3d 456, 470 (5th Cir. 2000) (Garza, J., concurring).

IV. Evidentiary Federalism and Due Process in Death Penalty Proceedings

Even if admitting clinical predictions into testimony violates rules of evidence, federal courts will not hear claims of state law violations unless the state law violations amount to violations of the federal Constitution.[91] Questions of evidence law in state courts are state law questions, and courts do not usually grant habeas relief for failure to follow the state evidentiary rules.[92]

---

91. *See* Estelle v. McGuire, 502 U.S. 62, 67-68 (1991) (finding that "it is not the province of a federal habeas court to reexamine state court determinations on state-law questions").

92. *See* Spencer v. Texas, 385 U.S. 554, 563-64 (1967) ("[T]he Due Process Clause guarantees the fundamental elements of fairness in a criminal trial. But it has never been thought that such cases establish this Court as a rule-making organ for the promulgation of state rules of criminal procedure." (citations omitted)). Moreover, evidentiary rules do not necessarily apply at sentencing hearings. Federal Rule of Evidence 1101(d)(3) states that the rules do not apply to sentencing proceedings, so federal courts do not apply them. The states are divided about their application. *See* Robert A. Kelly, *Applicability of the Rules of Evidence to the Capital Sentencing Proceeding: Theoretical & Practical Support for Open Admissibility of Mitigating Information*, 60 UMKC L. REV. 411, 457 (1992) (analyzing capital sentencing statutes and concluding that nineteen states do not apply the rules of evidence and that seventeen states use evidentiary rules for at least part of their sentencing proceedings, but that all thirty-six states require that any evidence admitted be relevant and probative). Nonetheless, all states invoking the death penalty provide that any evidence admitted be both relevant and material. *See id.* (stating that for evidence to be relevant and probative—as all death penalty states require—evidence must be "logically relevant, legally relevant, and offered for the purpose of proving or supporting a material proposition"); *see also* U.S. SENTENCING GUIDELINES MANUAL § 6A1.3 ("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 . . . ."); ALA. CODE § 13A-5-45(d) (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"). Hearsay, for example, is frequently admissible in sentencing hearings. *See, e.g.*, Williams v. New York, 337 U.S. 241, 251 (1949) (noting the different evidentiary standards for sentencing); Todd v. Schomig, 283 F.3d 842, 853 (7th Cir. 2002) (allowing hearsay evidence in sentencing phase). Testimony about unadjudicated prior offenses is also allowed. *See, e.g.*, United States v. Lee, 274 F.3d 485, 494 (8th Cir. 2001) (allowing evidence of unadjudicated prior offenses); Gilbert v. State, 951 P.2d 98, 122 (Okla. Crim. App. 1997) (same). 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, 163 (1992) (finding evidence presented at sentencing hearing that the convicted murderer was a member of the Aryan Brotherhood irrelevant); United States v. Huckins, 53 F.3d 276, 279–80 (9th Cir. 1995) (finding that hearsay, to be admissible at sentencing, must have some other corroboration). The floor of relevance and materiality, in the context of science, is that the testimony reflect valid science. Although the Supreme Court in *Daubert* addressed the admissibility of scientific evidence under the Federal Rules of Evidence, it based its reasoning on concepts of relevance and probative value, which it linked to the requirement that the evidence assist the jury. *See* Daubert v. Merrell Dow Pharms., 509 U.S. 579, 591 (1993) (stating that assisting the trier of fact "goes primarily to relevance").

In *Estelle v. McGuire*,[93] for example, the Supreme Court found that it had been error to grant habeas relief on the basis of admitted battered child syndrome evidence introduced to prove intent in the trial of a father for murdering his six-month-old daughter.[94] The U.S. Court of Appeals for the Ninth Circuit granted habeas based on its holding that this evidence was "incorrectly admitted . . . pursuant to California law"[95] and that in conjunction with an instruction about the use of prior act evidence, it amounted to a due process violation.[96] Because the prosecution proffered the evidence to show that the injuries were not accidental, but instead were a product of child abuse, the Supreme Court held that the evidence of prior injuries was relevant to show intent, even if it did not purport to show who caused the injuries.[97] The Court noted that "nothing in the Due Process Clause of the Fourteenth Amendment requires the State to refrain from introducing relevant evidence simply because the defense chooses not to contest the point" (the defense did not contest that somebody had abused the baby).[98]

The only evidentiary question that the habeas court may review is whether the evidence "so infected" the proceeding that it resulted in the violation of a constitutional right.[99] The Court held that neither the battered child syndrome evidence nor the instruction as to its use amounted to such a violation because the evidence was relevant to intent, served to narrow the possible perpetrators, and was consistent with "the familiar use of evidence of prior acts for the purpose of showing intent, identity, motive or plan."[100] Thus, the framework that the Court established is that a mere violation of state evidentiary rules will not be sufficient for federal relief; some constitutional violation must have occurred.[101]

The flip side to *McGuire*, however, is that the failure to meet the threshold standards of relevance and reliability in expert testimony is no mere state law failure, but a constitutional error. Notably, no one contested the scientific validity of the testimony in *McGuire*. Rather, the dispute was over its relevance for proving intent, and the Supreme Court found it to be relevant to that

93. Estelle v. McGuire, 502 U.S. 62 (1991).

94. *Id.* at 67.

95. *Id.* (quoting McGuire v. Estelle, 902 F.2d 749, 754 (9th Cir. 1990)).

96. *Id.* at 66–67.

97. *Id.* at 68.

98. *Id.* at 70.

99. *Id.* at 72.

100. *Id.* at 75 (citing FED. R. EVID. 404(b)).

101. *See* Marshall v. Lonberger, 459 U.S. 422, 438 n.6 (1983) ("[T]he Due Process Clause does not permit the federal courts to engage in a finely tuned review of the wisdom of state evidentiary rules . . . .").

issue. But the issue of expert future dangerousness testimony poses a relevance issue at a threshold due process level. That is, if the expert testimony about future dangerousness were scientifically valid, it would unquestionably be relevant to the reasoned moral inquiry about retribution for the defendant. Cross-examination could expose any weaknesses. But if the expert testimony has no basis in reality, if it is not grounded in science, then it is wholly irrelevant to the jury's task, and its admission is arbitrary.[102]

Imposing the ultimate sentence ought not to be arbitrary.[103] Throughout its death penalty jurisprudence, the Court has maintained that "there is a significant constitutional difference between the death penalty and lesser punishments."[104] In *Furman v. Georgia*,[105] the Supreme Court found that the imposition of the death penalty in three cases violated the Eighth and Fourteenth Amendments as "cruel and unusual punishment."[106] In a subsequent series of cases, the Court reinstated the death penalty, as long as the authorizing legislation guided discretion in its imposition.[107]

In its more recent cases, the Supreme Court has defined two important Eighth Amendment principles in its death penalty jurisprudence: individualized sentencing and eliminating arbitrariness through channeled discretion.[108] The "evenhanded, rational, and consistent imposition of death sentences under

---

102. *Cf.* Logerquist v. McVey, 1 P.3d 113, 115 (Ariz. 2000) (finding that although neither *Daubert* nor *Frye* applied to the admissibility of expert testimony, the offering party still must establish scientific validity for the evidence to meet the threshold reliability and relevance standards).

103. *See* Gregg v. Georgia, 428 U.S. 153, 189 (1976) (explaining that "discretion must be suitably directed and limited so as to minimize the risk of wholly arbitrary and capricious action"); *see also* Jurek v. Texas, 428 U.S. 262, 276 (1976) (finding that a Texas statutory scheme for imposing the death penalty is rational and therefore constitutional).

104. Beck v. Alabama, 447 U.S. 625, 637 (1980) (plurality opinion).

105. Furman v. Georgia, 408 U.S. 238 (1972).

106. *Id.* at 239–40.

107. *See Gregg*, 428 U.S. at 189 (stating that "discretion must be suitably directed and limited so as to minimize the risk of wholly arbitrary and capricious action"); *Jurek*, 428 U.S. at 276 (finding that a state statutory scheme for imposing the death penalty is rational and therefore constitutional); Proffitt v. Florida, 428 U.S. 242, 259–60 (1976) (same).

108. *See* Lockett v. Ohio, 438 U.S. 586, 604 (1978) (prohibiting the preclusion of, "as a mitigating factor, any aspect of a defendant's character or record and any of the circumstances of the offense that the defendant proffers as a basis for a sentence less than death"); *see also* Kirchmeier, *supra* note 21, at 346 (observing that the Supreme Court, in its struggle to create a fair system, "has rejected both mandatory death penalty schemes and schemes that give total unbridled discretion to the sentencer"). *But see* Walton v. Arizona, 497 U.S. 639, 664 (1990) (Scalia, J., concurring) (advocating a mandatory death penalty scheme because "to refer to the two lines as pursuing 'twin objectives,' . . . is rather like referring to the twin objectives of good and evil. They cannot be reconciled." (quoting Spaziano v. Florida, 468 U.S. 447, 459 (1984))).

law" was the goal.[109] In making its decision, the sentencing body must reach "a reasoned moral response" free of impediments to relevant sentencing considerations.[110] Expert testimony that purports to do what it cannot do would appear to be precisely such an impediment.

Not only the Eighth Amendment, but also the Due Process Clause of the Fourteenth Amendment mandates that the sentencing body make determinations in a manner that is not "arbitrary and capricious."[111] Because many of the rights given at trial do not apply at sentencing, due process is the primary source of regulation.[112] Due process governs the right to accurate information.[113] In *Gregg v. Georgia*,[114] 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."[115] Moreover, because of the severity of the sanction, courts must handle the imposition of death in a way that permits a defendant to present whatever facts might impel a jury to mercy.[116]

Fundamental fairness, the freestanding content of the Fourteenth Amendment, poses the question of whether the procedure subjects the defendant to an

---

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

110. California v. Brown, 479 U.S. 538, 545 (1987).

111. *See* Gregg v. Georgia, 428 U.S. 153, 189 (1976) ("[W]here discretion is afforded a sentencing body on a matter so grave as the determination of whether a human life should be taken or spared, that discretion must be suitably directed and limited so as to minimize the risk of wholly arbitrary and capricious action."). This due process right, sometimes called "free-standing due process" because it stands apart from the Bill of Rights, is a procedure "necessary to an Anglo-American regime of ordered liberty." Duncan v. Louisiana, 391 U.S. 145, 159 n.14 (1968). For a discussion of the jurisprudence of free-standing due process in the Supreme Court, see generally Jerold H. Israel, *Free-Standing Due Process and Criminal Procedure: The Supreme Court's Search for Interpretive Guidelines*, 45 ST. LOUIS U. L.J. 303 (2001).

112. *See* Israel, *supra* note 111, at 393 (observing that at "the sentencing stage, most trial-type rights (e.g., confrontation) do not apply and due process becomes the primary source of constitutional regulation").

113. *See* Williams v. New York, 337 U.S. 241, 245–52 (1949) (finding that due process allows a judge to consider out-of-court sources of information during sentencing).

114. Gregg v. Georgia, 428 U.S. 153 (1976) (plurality opinion).

115. *Id.* at 190. Along with its two companion cases, *Proffitt v. Florida*, 428 U.S. 242 (1976) (plurality opinion), and *Jurek v. Texas*, 428 U.S. 262 (1976) (plurality opinion), the Supreme Court authorized the states to return to capital punishment in *Gregg*.

116. *See* Lockett v. Ohio, 438 U.S. 586, 605 (1978) ("Given that the imposition of death by public authority is so profoundly different from all other penalties, we cannot avoid the conclusion that an individualized decision is essential in capital cases . . . [and that] treating each defendant in a capital case with that degree of respect due the uniqueness of the individual is far more important than in noncapital cases.").

unacceptably high risk of erroneous decisions.[117] In the criminal context, however, due process proscribes only those state procedures that "offend[] some principle of justice so rooted in the traditions and conscience of our people as to be ranked as fundamental."[118] Thus, in determining whether the future dangerousness testimony implicated a fundamental right, the Court has looked to principles having "deep roots" in our heritage.[119]

A requirement of "deep roots" does not limit due process to historical practices, however. Instead, it requires adherence to basic rule-of-law principles. For example, in rejecting a state position that put the decision about lineup reliability to the jury, the Supreme Court held that a substantial likelihood of misidentification precluded the admission of lineup identification evidence as a violation of due process, even without reference to common law traditions or contemporary consensus.[120] Two major due process concerns are apparent with respect to expert testimony in capital sentencing: it really matters that the facts presented by experts are accurate, and the jury should have access to all of the accurate facts that might help it reach a reasoned moral decision.[121]

The Supreme Court's death penalty jurisprudence since *Furman v. Georgia* has focused on balancing consistency in decisionmaking with individual

---

117. *See, e.g.*, Crane v. Kentucky, 476 U.S. 683, 690 (1986) (finding that due process and compulsory process analyses require the admission of reliable exculpatory defense evidence regarding the credibility of a confession); United States v. Bagley, 473 U.S. 667, 669 (1985) (stating that disclosure of exculpatory evidence is required under due process (citing Brady v. Maryland, 373 U.S. 83, 87 (1963))); Ake v. Oklahoma, 470 U.S. 68, 84 (1985) (stating that due process requires a state to provide an expert for the defense under certain circumstances); Foster v. California, 394 U.S. 440, 443 (1969) (excluding suggestive lineup on due process grounds); Rideau v. Louisiana, 373 U.S. 723, 726 (1963) (upholding a constitutional due process right to a change in venue to protect the defendant from local prejudice); Napue v. Illinois, 360 U.S. 264, 269 (1959) (finding that due process prohibits the prosecution's knowing use of perjured testimony); *see also* Israel, *supra* note 111, at 386 (discussing the role of subjectivity in the application of the fundamental fairness standard).

118. Medina v. California, 505 U.S. 437, 445 (1992) (quoting Patterson v. New Jersey, 432 U.S. 197, 201–02 (1977)).

119. *Id.* at 446 ("The rule that a criminal defendant who is incompetent should not be required to stand trial has deep roots in our common-law heritage.").

120. In three important cases decided on the same day, the Supreme Court determined that eyewitness identification is unreliable and therefore that the right to counsel attached. The cases were *Stovall v. Denno*, 388 U.S. 293 (1967); *Gilbert v. California*, 388 U.S. 268 (1967); and *United States v. Wade*, 388 U.S. 218 (1967). *See* Israel, *supra* note 111, at 417 (noting a number of Supreme Court free-standing due process decisions that addressed neither historical practice nor contemporary consensus).

121. *See Lockett*, 438 U.S. at 605 (discussing the need for individualized sentences in capital cases).

fairness.[122] The Supreme Court set the minimal standard for imposing the death penalty as "the evenhanded, rational, and consistent imposition of death sentences under law."[123] Practices that threaten the reliability of the sentencing procedure cannot fall within such a framework. Permitting experts to testify without scientific bases for their assertions is thus inimical to due process standards. Admissibility of scientific evidence should depend upon its relevance and reliability. The unresolvable problem of clinical predictions of future dangerousness is that they have no scientific basis. Admitting expert testimony that has no scientific basis is misleading to the jury and thus precludes rational decisionmaking.

If a state's procedure precludes reliable, rational decisionmaking, it violates due process.[124] The Due Process Clause of the Fourteenth Amendment provides a right to fundamentally fair proceedings, and a relevance requirement is an integral part of a fair and rational proceeding.[125] The requirement of fundamental fairness encompasses a requirement of rationality, so that the state may not present the jury with inaccurate or misleading evidence.[126] Because

---

122. *See* Jonathan R. Sorensen & Rocky L. Pilgrim, *An Actuarial Risk Assessment of Violence Posed by Capital Murder Defendants*, 90 J. CRIM. L. & CRIMINOLOGY 1251, 1251 (2000) ("[D]eath penalty jurisprudence has focused on how to insure consistency in decisionmaking while providing fairness to individual defendants.").

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

124. Notably, the Supreme Court announced a three-part balancing test in *Mathews v. Eldridge*, 424 U.S. 319 (1976): consideration of the private interest that state action would affect, risk of erroneously depriving someone of that interest through the procedures used, and the government's interest (including fiscal and administrative burdens imposed). *Id.* at 334–35. In *Medina v. California*, however, it declined to apply such a balancing test to state procedural rules that are part of criminal procedure because it believed that any expansion of explicit constitutional guarantees "invites undue interference with both considered legislative judgments and the careful balance the Constitution strikes between liberty and order." Medina v. California, 505 U.S. 437, 443 (1992). However, as Jerold Israel explains, *Mathews*'s concept of freestanding due process has purpose not so much in balancing, but in looking at the "logical implications of a basic principle of fairness." Israel, *supra* note 111, at 423. Permitting expert testimony that borders on the fraudulent can do little other than undermine these basic principles. Israel suggests that one can interpret the Court's fundamental fairness jurisprudence in four ways: (1) insistence on "a few basic elements of trial-type adjudication" with the addition of "a wide variety of rational procedures"; (2) assuring that the process does not convict the innocent; (3) giving the specific guarantees of the Bill of Rights preemptive influence; (4) giving the basic procedural protection of the common law with the addition of a cost-benefit analysis. *Id.* at 425.

125. *See* Payne v. Tennessee, 501 U.S. 808, 825 (1991) (stating that the Due Process Clause provides a right to relief when a proceeding is "fundamentally unfair"); Lankford v. Idaho, 500 U.S. 110, 121 (1991) (finding that the Clause represents a profound attitude of fairness between the individual and the government).

126. *See* United States v. Scheffer, 523 U.S. 303, 308 (1998) (upholding military rule

testimony that is without scientific merit is inimical to rational decisionmaking, a state process permitting such testimony undermines the fundamental fairness of the proceeding.[127]

The due process concern was paramount in the Supreme Court's ruling in *Simmons v. South Carolina*,[128] in which the Court held that the sentencing jury was entitled to be accurately informed of their state's alternative to the death penalty.[129] The *Simmons* Court acknowledged the primacy of future dangerousness to sentencing decisions and emphasized how the lack of information about eligibility for parole could skew the sentencing decision.[130] This primacy of future dangerousness to the sentencing determina-tion—ninety-five percent of death sentencing jurors believe that the evidence showed that the defendant posed a threat to society[131]—similarly mandates that expert testimony proffered to establish future dangerousness must be accurate.[132] The premise that the jury is entitled to accurate information in reaching its decision would appear counter to the Court's refusal to exclude misleading testimony in *Barefoot*. If expert testimony is inaccurate, as it is in clinical predictions of future dangerousness, then it does not meet the indispensable prerequisite of providing the jury with accurate information.

---

excluding polygraph evidence as a rational interest in eliminating unreliable evidence); Caldwell v. Mississippi, 472 U.S. 320, 340 (1985) (finding the prosecution's argument at sentencing that a juror's decision was reviewable unconstitutional because it was both inaccurate and misleading); Chambers v. Mississippi, 410 U.S. 284, 300 (1973) (holding hearsay evidence admissible because the statements had "considerable assurance of their reliability"); Washington v. Texas, 388 U.S. 14, 22 (1967) (addressing the defendant's right to present witnesses and finding a rule of exclusion arbitrary if it had no rational basis).

127. Fundamental fairness is infrequently invoked by the Supreme Court and, when invoked, is generally without explanation. *See, e.g.*, Payne v. Tennessee, 501 U.S. 808, 825 (1991) (noting that the Due Process Clause provides relief when the introduction of evidence—there, victim impact evidence—makes the trial "fundamentally unfair"); Lassiter v. Dept. of Soc. Servs., 452 U.S. 18, 24 (1981) (fundamental fairness is "a requirement whose meaning can be as opaque as its importance is lofty").

128. Simmons v. South Carolina, 512 U.S. 154 (1994) (plurality opinion).

129. *See id.* at 169 (observing that "parole was a mainstay" in most states and that "most juries lack accurate information about the precise meaning of 'life imprisonment'").

130. *Id.* at 163 (noting that "it is entirely reasonable for a sentencing jury to view a defendant who is eligible for parole as a greater threat to society than a defendant who is not").

131. *See* Theodore Eisenberg & Martin T. Wells, *Deadly Confusion: Juror Instructions in Capital Cases*, 79 CORNELL L. REV. 1, 6 (1993) (cited by the plurality opinion in *Simmons*).

132. *See, e.g.*, Stephen P. Garvey, *Aggravation and Mitigation in Capital Cases: What Do Jurors Think?*, 98 COLUM. L. REV. 1538, 1560 (1998) (reporting on research data demonstrating the "pervasive role future dangerousness plays in and on the minds of capital sentencing jurors").

V. Gatekeeping and Capital Sentencing

A. Capital Sentencing Proceedings

The excuse for the courts' refusal to scrutinize expert testimony at sentencing is that the rules of evidence do not generally apply. The reason that the rules of evidence do not generally apply at sentencing proceedings is that sentencing proceedings historically are nonadversarial.[133] In the pre-sentencing guidelines era, the judge had access to a wide variety of information and had discretion in imposing sentences within statutory prescriptions. Although the federal and states' guideline regimes have narrowed judicial discretion somewhat, the norm remains judicial access to a wide variety of background material. This structure makes ordinary sentencing similar to the European inquisitorial system of adjudication, in which the judge has access to the defendant's file, both parties may present their views, but only the judge directs the investigation and questions the witnesses.[134] Under such a regime, little concern for rules of evidence exists—the judge is presumed to rely only on relevant information. The idea of gatekeeping is therefore unnecessary, and the continental view of relevance is very technical and precise.[135] On the other hand, in the adversarial mode, it is the parties who control, initially at least, what information will be available to decisionmakers, and judicial gatekeeping ensures that the proceedings do not degenerate into a show that will mislead the jury from its mission and that the details that the litigants present have some propensity to bring out the truth.[136]

Capital sentencing is different from other criminal sentencing. Following a constitutionally mandated bifurcated trial,[137] a jury performs the capital sentencing.[138] A common justification for declining to apply the rules of

---

133. *Supra* note 92.

134. *See* Damaska, *supra* note 71, at 1083–1106 (contrasting the adversarial and inquisitorial systems of adjudication). In an inquisitorial system, the prosecution and the defense do not have separate witnesses, and it is the judge's responsibility to question the witnesses and inquire into the record. *Id.* Of course, in practice, there are no purely inquisitorial or purely adversarial systems; both borrow from each other. *See* MIRJAN DAMASKA, THE FACES OF JUSTICE AND STATE AUTHORITY: A COMPARATIVE APPROACH TO THE LEGAL PROCESS 241 (1986) (explaining that as "applied to existing systems, most of them were found to be pastiches").

135. *See* Damaska, *supra* note 71, at 1104–05 ("The continental system would tend to embrace a paradigm closer to that of scientific investigation.").

136. *See id.* at 1090–91 (describing the adversarial mode of proof-taking).

137. Lockett v. Ohio, 438 U.S. 586, 605 (1978) (mandating bifurcated trials to accommodate the need for both guided discretion and "individualized consideration" in capital sentencing).

138. *See* Ring v. Arizona, 536 U.S. 584, 589 (2002) (ruling that the capital sentencing decision must be made by a jury).

evidence strictly at sentencing is that such proceedings have become a search for justice rather than truth. As the Supreme Court held in *Lockett v. Ohio*,[139] the evidentiary standards in the sentencing phase are fairly open because "any aspect [of the defendant's] character or record and any of the circumstances of the offense" ought to be available to support a sentence less than death.[140] The importance of accuracy, however, in the context of what is an unabashedly adversarial proceeding, demands some control over what counts as facts in the proceeding.

Thus, the jury's "reasoned moral response" to the evidence and arguments at sentencing, according to the Supreme Court, must have a basis in information sufficient and relevant for reliable, rational decisionmaking.[141] At capital sentencing hearings, both prosecution and defense present testimony and argument. Expert testimony is prevalent.[142] Without formal requirements of evidentiary rules, virtually no structured examination of the scientific basis for such testimony exists. Despite its adversarial character, judges frequently believe that they are not bound to exercise gatekeeping duties. As a result, the courts exercise gatekeeping functions in sporadic and apparently arbitrary ways.

B. Future Dangerousness Testimony in Capital Sentencing

Both state and federal courts are lackadaisical about gatekeeping when it comes to capital sentencing.[143] The Federal Rules of Evidence—which

---

139. Lockett v. Ohio, 438 U.S. 586 (1978) (plurality opinion).

140. *Id.* at 604 (plurality opinion).

141. Penry v. Lynaugh, 492 U.S. 302, 319 (1989) (explaining that "the sentence imposed at the penalty stage should reflect a reasoned moral response to the defendant's background, character, and crime" (quoting California v. Brown, 479 U.S. 538, 545 (1987))).

142. Parties frequently offer expert testimony at capital sentencing proceedings. For example, in the Capital Jury Project, funded by the National Science Foundation, the California portion of the study examined thirty-six death penalty cases and found that the prosecution called an expert in eighty-one percent of the cases, and the defense called an expert in ninety percent. Scott E. Sundby, *The Jury as Critic: An Empirical Look at How Capital Juries Perceive Expert and Lay Testimony*, 83 VA. L. REV. 1109, 1120 (1997) (noting that "conventional practice at the penalty phase involves presenting an expert to the jury at some point—in many cases more than one—who will testify based upon an expertise gained through training and study").

143. The Violent Crime Control and Law Enforcement Act of 1994 significantly expanded the scope of federal capital crimes. Violent Crime Control and Law Enforcement Act of 1994, Pub. L. No. 103-322, 108 Stat. 1796 (codified in scattered sections of 18 U.S.C.); *see* Charles K. Eldred, *Recent Developments, The New Federal Death Penalties*, 22 AM. J. CRIM. L. 293, 296–98 (1994) (listing sixty capital offenses). *But see* Robert F. Worth, *Prosecutors Oppose Judge in Ruling on Death Penalty*, N.Y. TIMES, May 27, 2002, at B3 (citing opinion by Judge Jed Rakoff stating that the death penalty posed "an undue risk that a meaningful number of

include the *Daubert* standard for expert witnesses—do not apply at federal capital sentencing proceedings,[144] despite the Supreme Court's recognition that capital sentencing requires a "heightened reliability" standard.[145] Nonetheless, by statute, only relevant evidence is admissible,[146] and as under the Federal Rules of Evidence, if the danger of unfair prejudice outweighs its probative value, courts may exclude even relevant evidence.[147] Thus, even if the federal rules do not apply directly, some scientific validity inquiry ought to apply.

The courts occasionally recognize this tension between heightened reliability for death sentencing proceedings and the inapplicability of evidentiary rules by at least referring to a scientific validity inquiry, even though the court declines to reach the *Daubert* issue. For example, in *United States v. Barnette*,[148] the court disposed of prosecution arguments that the federal rules do not apply at sentencing and found that the contested future dangerousness evidence met *Daubert* standards.[149] The defense argued that prosecution psychiatric testimony predicting dangerousness based on an actuarial instrument, the Hare Psychopathy Checklist-Revised (PCL-R), did not meet *Daubert* standards of reliable scientific evidence.[150] The defense argued that three of the PCL-R checklist factors were impermissible: race, poverty, and

---

innocent persons" were being executed); Thomas F. Liotti, *Criminal Justice System in Crisis*, N.Y.L.J., Jan. 21, 2003, at S5 (noting "Judge Jed Rakoff's courageous recent decision declaring the federal death penalty unconstitutional"). Even in federal cases, however, courts do not generally invoke *Daubert* at sentencing because the rules of evidence do not apply. *See, e.g.*, U.S. SENTENCING GUIDELINES MANUAL § 6A1.3(a) (2002) ("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 . . . ."); ALA. CODE § 13A-5-45(d) (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").

144. *See* 18 U.S.C. § 3593(c) (2000) ("[I]nformation is admissible regardless of its admissibility under the rules governing admission of evidence at criminal trials.").

145. *See* 21 U.S.C. § 848(j) (2000) (allowing the presentation of aggravating and mitigating factors at sentencing); Ford v. Wainwright, 477 U.S. 399, 411 (1986) (setting a "heightened standard of reliability" for capital sentencing).

146. *See* 18 U.S.C. § 3593(c) (2000) (stating that "information may be presented as to any matter relevant to the sentence").

147. *Id.* § 3593(a) (analogous to FED. R. EVID. 403).

148. United States v. Barnette, 211 F.3d 803 (4th Cir. 2000).

149. *Id.* at 815–16 (concluding, in response to government arguments that "*Daubert* does not apply to the sentencing phase of a capital trial," that the court "need not address whether *Daubert* applies to sentencing hearings, because, even assuming that it does, we find the evidence meets its standards for admissibility").

150. *Id.*

age.[151] The Fourth Circuit Court of Appeals noted that the defense did not contest relevancy and found that the defense objections to reliability were based on only two publications by the defense expert.[152] The court held that this basis was insufficient to overturn the trial judge's reliability determination.[153] The court distinguished the holding in *United States v. Powers*[154] because there the proponent of the testimony could not rebut the showing of unreliability,[155] and in contrast, "Barnette only proffered two articles written by his expert."[156] Although the court acknowledged that the PCL-R did use the three impermissible factors, the court found that the expert had not relied on them exclusively, but also on seventeen other PCL-R factors, the Diagnostic and Statistical Manual, Fourth Edition (DSM-IV), personal observations, and research on predicting future dangerousness.[157]

Similarly, in *United States v. Lee*,[158] the U.S. Court of Appeals for the Eighth Circuit upheld the admissibility of the prosecution expert's psychological testimony that the defendant showed no remorse, without discussing scientific validity.[159] In *Lee*, the court held that the defense had opened the door by presenting psychological testimony in mitigation.[160] Although the prosecution expert exceeded the scope of direct testimony, the court held that the statement "fell within the wide boundaries set for the admission of evidence at capital sentencing hearings" and was therefore admissible.[161] Worse, the defense's psychological expert opened the door to testimony about unadjudicated prior offenses, such as various assaults, which the court not only held were probative of future dangerousness, but also impeached the testimony of the defense expert.[162] As a result of these findings, the Eighth Circuit overturned the district court's order for a new penalty phase hearing.[163]

---

151. *Id.*

152. *Id.*

153. *See id.* (finding that the trial court's examination of the issue did not support a finding of abuse of discretion).

154. United States v. Powers, 59 F.3d 1460, 1471 (4th Cir. 1995) (holding that the penile plethysmograph test failed to meet scientific validity standards and was therefore inadmissible).

155. *Id.*

156. United States v. Barnette, 211 F.3d 803, 816 (4th Cir. 2000).

157. *Id.*

158. United States v. Lee, 274 F.3d 485 (8th Cir. 2001).

159. *See id.* at 492–95 (discussing admissibility of expert psychological testimony).

160. *Id.* at 495.

161. *Id.*

162. *See id.* ("By introducing a mental health expert in defense, Lee opened the door to testimony concerning psychological diagnosis.").

163. *Id.* at 497.

One finds a third example of how the basic evidentiary precepts may become distorted in death penalty sentencing in *United States v. Allen*.[164] There, the Eighth Circuit found that the Federal Death Penalty Act, which permits "information" rather than "evidence" to be admitted at the sentencing phase, did not violate the Eighth Amendment as inherently unreliable.[165] "[N]ot only does the statute protect a defendant from both irrelevant information and overly prejudicial information, the relaxed evidentiary standard also works to a defendant's advantage in helping to prove mitigating factors and to disprove aggravating factors."[166] At issue in the case, however, was the prosecution's reference to the defendant as a "murderous dog" in arguing future dangerousness at closing.[167] The court found this argument to be improper but not so unduly prejudicial as to violate the defendant's due process rights to fair sentencing.[168]

Even without expert predictions, future dangerousness may become an issue when the prosecution attempts to demonstrate future dangerousness by showing that the defendant has a past criminal history.[169] This past conduct frequently comes into evidence at sentencing through unadjudicated prior crimes.[170] Although a pattern of previous violence is one of the strongest predictors of future violence,[171] the evidence of such conduct is often flimsy

---

164. United States v. Allen, 247 F.3d 741 (8th Cir. 2001), *cert. granted and judgment vacated and remanded in light of* Ring v. Arizona, 122 S. Ct. 2653 (2002).

165. *Id.* at 759.

166. *Id.*

167. *Id.* at 775.

168. *Id.* at 777.

169. *See, e.g.*, Gilliard v. Scroggy, 847 F.2d 1141, 1147 (5th Cir. 1988) (finding that a statutory aggravating factor allowing the jury to consider past convictions for violent offenses "implicitly" posed a future dangerousness inquiry).

170. It is a common practice at sentencing for the prosecution to introduce evidence that the defendant committed other crimes that were never tried. *See* Steven Paul Smith, Note, *Unreliable and Prejudicial: The Use of Extraneous Unadjudicated Offenses in the Penalty Phase of Capital Trials*, 93 COLUM. L. REV. 1249, 1250 (1993) (observing that "states have adopted diverse and inconsistent approaches" and that the circuit courts disagree on the constitutionality of this evidence). The Supreme Court addressed the constitutionality of basing the sentence on a presentencing investigative report without disclosing it to the defendant in *Gardner v. Florida*, 430 U.S. 349 (1977). Because the defendant had no opportunity to deny or explain the contents, the Court held the evidence to be so unduly prejudicial that it rendered the trial fundamentally unfair. *Id.* at 362.

171. *See* BENCHBOOK, *supra* note 34, at 223 ("A history of past violence repeatedly has been shown to be one of the best predictors of violence."); Deidre Klassen & William A. O'Connor, *A Prospective Study of Predictors of Violence in Adult Male Mental Health Admissions*, 12 LAW & HUM. BEHAV. 143, 151 tbl.1 (1988) (finding that a prior history of violent crime associated strongly with post-release arrests in studies of men released from

at best, and jurors fallaciously equate crime with violence.[172] In response to the unreliability of such evidence, some courts require a showing of sufficient indicia of reliability so that introducing the evidence does not make the sentencing hearing fundamentally unfair.

For example, in considering the constitutional implications of admitting evidence of prior unadjudicated offenses to prove future dangerousness, the U.S. District Court for the Eastern District of Virginia recognized a heightened reliability standard for capital sentencing and stated that the court, although not bound to observe the rules of evidence regarding admissibility, should not permit "an evidentiary free-for-all that undermines reliability."[173] 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.[174] Courts should require a similar approach to expert testimony—a *Daubert*-like hearing outside the presence of the jury to determine scientific validity of the expert testimony—to meet constitutional demands for relevance and reliability of future violence predictions.

C. Gatekeeping, Epistemic Norms, and Jury Decisionmaking

Apart from the constitutional and rule-of-law due process concerns, there remains the question of why we need gatekeepers. Epistemology—the study of the acquisition of knowledge—has important bearing on this question. Evidence rules, or the lack of them, determine what knowledge the jury will be able to use in its determination.[175] From an epistemic vantage point, the question of whether we need judges to act as gatekeepers focuses on whether

---

psychiatric hospital).

172. Notably, a single past violent act is not enough to form a pattern. Moreover, relying on past criminal charges poses a severe problem both with respect to false positives (the charge may not reflect actual conduct but only reports of conduct) and false negatives (charges may be absent in the presence of some behaviors such as stalking and domestic violence). *See, e.g.*, Paul H. Robinson, *Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114 HARV. L. REV. 1429, 1430 (2001) (arguing against three-strikes laws). Moreover, although a pattern of violence is predictive of future violence, a pattern of crime is not.

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

174. *Id.* at 1000.

175. Brian Leiter, *The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make for Good Philosophy of Evidence*, 1997 BYU L. REV. 803, 805 (explaining that if *Frye* is wrong, it is for reasons of law, not reasons of science, because the epistemic norms in science and law are different). I agree with Professor Leiter that the rules of admissibility need to be evaluated in terms of the goals of law rather than the goals of science, but I disagree with his conclusion that jurors are as competent as judges in evaluating junk science. Beecher-Monas, *supra* note 13, at 75.

gatekeeping actually promotes the acquisition of knowledge in the context of a particular social practice, here the capital sentencing hearing.[176]

The capital jury makes reasoned moral judgments about the fate of the defendant—a policy decision.[177] In order to reach a just decision, it is supposed to be provided with "information relevant to the imposition of the sentence and provided with standards to guide its use of the information."[178] Truth—accuracy—is a goal of legal policy decisions as well as legal fact determinations. The purpose of permitting experts—witnesses without personal knowledge of the defendant or incident—serves the same purpose in capital sentencing as at trial: to inform the jury about matters outside their common experience in order to effectuate the truth-seeking goals of the rule of law. The question is therefore whether initial screening by the judge for scientific validity aids in this task. The Supreme Court in *Barefoot* found that society could rely upon the adversary system to present enough information to jurors so that they could sort reliable from unreliable expert testimony. The *Daubert* Court found that courts needed to screen expert testimony for relevance first. Which was correct? Two lines of research about juror decisionmaking suggest a framework for analysis.

1. The Story Model

A model of juror decisionmaking proposed by Pennington and Hastie is the story model, in which the juror's decision depends on how well the expert's testimony fits with the juror's pre-existing views.[179] Although Pennington and Hastie did not directly apply their framework to expert testimony, they posited a model consisting of three parts: first, the jurors use their own preconceptions to weave a story from the evidence that they heard at trial; second, they take the jury instructions and create verdict alternatives; third, they attempt to find the best correlation between the story that they con-

---

176. The question asked by social epistemology is what norms work best under the real world limits of a particular social practice. *See* Leiter, *supra* note 175, at 814–15 (suggesting two lines of inquiry: "paternalism," whether substituting judicial screening will enable jurors to make more accurate decisions; and "ought equals can," that is, whether shortcomings in cognition will preclude either judges or jurors from making an accurate decision).

177. *See* California v. Ramos, 463 U.S. 992, 1008 (1983) (stating that the jury, at the penalty phase, no longer acts strictly as fact finder; rather, the jury's mission involves "a myriad of factors"); Ford v. Strickland, 696 F.2d 804, 831 n.17 (11th Cir. 1983) (Tjoflat, J., concurring) ("The sentencer, therefore, acts not as fact finder, but as policymaker.").

178. Gregg v. Georgia, 428 U.S. 153, 195 (1976).

179. *See* Nancy Pennington & Reid Hastie, *Evidence Evaluation in Complex Decision Making*, 51 J. PERSONALITY & SOC. PSYCHOL. 242, 243-45 (1986) (describing the story model of decisionmaking). *See generally* REID HASTIE ET AL., INSIDE THE JURY (1983).

structed and the verdict alternatives.[180] This model has a number of consequences regarding jury processing of expert testimony regarding future dangerousness.

Interviews with death penalty jurors demonstrate that connecting the facts of the case to the proffered expert opinion increases the influence of the expert opinion on the jury.[181] Thus, when an expert opines with certainty in response to the prosecutor's hypothetical reiteration of the facts of the case, it is very persuasive story-telling, even though it is the most inaccurate form of prediction. Not only is the opinion persuasive because of the story-telling aspect of the hypothetical, but its credibility increases with the strength of expert certainty. A number of surveys have shown that jurors rated experts who conveyed low confidence in their opinions as a cause for concern.[182] Thus, the confidence with which an expert offers his opinion bears a direct relationship to its credibility with the jury. Evidence that corresponds to strongly held beliefs is particularly persuasive.[183] For juries already predisposed to believe that a defendant poses a danger to society, a tendency to overvalue predictions that confirm such beliefs is a likely outcome.[184]

The adversary process exacerbates these problems. First, in experimental settings, expert clinical testimony was more persuasive to jurors than actuarial testimony, even after adversarial manipulations such as effective cross-exami-

---

180. *See* Pennington & Hastie, *supra* note 179, at 243-45 (explaining the stages of story construction, verdict category establishment, and story classification). The story model is similar to the idea of schemata in psychology. *See, e.g.*, David E. Rumelhart, *Schemata and the Cognitive System*, *in* 1 HANDBOOK OF SOCIAL COGNITION 161, 163 (Robert S. Wyer, Jr. & Thomas K. Srull eds., 1984). A schema is an "informal, private, unarticulated theory about the nature of the events, objects, or situations which we face. The total set of schemata we have available for interpreting our world in a sense constitutes our private theory of the nature of reality." *Id.* at 166.

181. Sundby, *supra* note 142, at 1144 (finding that jurors rated as more influential those experts who connected the facts of the case to their opinion).

182. *See* D. Shuman et al., *An Empirical Examination of the Use of Expert Witnesses in the Courts—Part II: A Three-Part Study*, 34 JURIMETRICS J. 193, 201 (1994) (finding that 14% of unfavorable expert witnesses did not appear very knowledgeable).

183. Certainty in witness testimony is especially persuasive when it correlates with strongly held beliefs. For example, jurors are likely to believe eyewitnesses who testify with certainty, and jurors are likely to disbelieve expert testimony with an inverse correlation between certainty and correctness, because jurors who strongly believe that witness confidence in identification correlates to accuracy tend to discount or ignore the expert. *See* Krauss & Sales, *supra* note 16, at 276 (citing studies); *see also* Sundby, *supra* note 142, at 1133 (citing interviews with capital jurors who found expert testimony about eyewitness evidence incredible because it contradicted their preconceptions).

184. *See* C. Walter Showalter & Richard J. Bonnie, *Psychiatrists and Capital Sentencing: Risks and Responsibilities in a Unique Legal Setting*, 12 BULL. AM. ACAD. PSYCH. & L. 159, 165 (1984) (jurors tend to overvalue predictions that confirm their beliefs).

nation and the testimony of a competing expert.[185] In addition, although effective cross-examination caused these jurors to re-evaluate the expert testimony, it had far less effect on clinical opinion testimony than on actuarial testimony.[186] This result may be because of juror familiarity with (and trust in) medical diagnoses for common diseases. These clinical predictions of violence differ from medical diagnoses, however. The best medical diagnosis requires personal examination and patient history, analyzed against a background of empirical data about the etiology of the disease that the physician is diagnosing. These factors are notably absent from clinical predictions of violence.[187]

2. Persuasion Theory and the Impact of Cognitive Quirks

The persuasion theory posits that when an argument is complex or difficult to understand, people rely on cognitive shortcuts to evaluate the argument and that expert qualifications then become a surrogate for trustworthiness.[188] Studies have shown that juror reliance on expert credentials is

---

185. *See* Krauss & Sales, *supra* note 16, at 291 (discussing the experiment and its results).

186. *Id.* at 302, 305 (observing that "[a]dversary procedures failed to return mock jurors who received clinical opinion expert testimony to their initial dangerousness rating levels" and concluding that jurors have a "predilection for less accurate clinical opinion testimony").

187. 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. *E.g.*, Charles W. Lidz et al., *The Accuracy of Predictions of Violence to Others*, 269 JAMA 1007 (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. *Id.* Random predictions would have a sensitivity and specificity of 50%. *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." 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 correctly reports that a person does not have the characteristic under investigation." *Id.* at 34. Actuarial studies, although more accurate than clinical predictions, still predict with less than stellar accuracy: when participants in 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).

188. *See* Regina A. Schuller & Neil Vidmar, *Battered Wife Syndrome Evidence in the Courtroom: A Review of the Literature*, 16 LAW & HUM. BEHAV. 273, 284 (1992) (stating that 80% of jurors found expert testimony influential).

directly proportional to the complexity of the information presented.[189] Moreover, there is some evidence that jurors tend to rank medical expertise higher than "merely" scientific expertise so that even when the information is identical, jurors listening to two experts, one a medical doctor and one a psychologist, find the medical expert more persuasive even though the testimony is identical.[190] Nothing suggests that judges are any different from jurors in this respect, but the kind of structured analysis required under *Daubert* counters this tendency by requiring the judge to evaluate the information presented.

In addition, there are a number of cognitive quirks—common to everyone, not just jurors—that may influence the capital sentencing decision. These quirks result from shortcuts that people use in making judgments in certain contexts.[191] People using these heuristics may be behaving rationally in the sense of conserving time, but forecasts differ from the rational actor approach of economics.[192] Judges also are prone to use these unconscious shortcuts, but a number of factors counteract this tendency in judges. First, the structured reasoning process undertaken in a *Daubert* inquiry improves performance in cognitive tasks.[193] Second, judges are repeat players in a way

---

189. Joel Cooper et al., *Complex Scientific Testimony: How Do Jurors Make Decisions?*, 20 LAW & HUM. BEHAV. 379, 379 (1996).

190. J. Greenberg & A. Wursten, *The Psychologist and the Psychiatrist as Expert Witnesses: Perceived Credibility and Influence*, 19 PROF. PSYCHOL. RES. & PRAC. 373, 378 (1988) (simulated insanity trial).

191. As Professor Mitchell points out, some people exhibit these quirks more than others and in some contexts more than others. *See* Gregory Mitchell, *Why Law and Economics' Perfect Rationality Should Not Be Traded for Behavioral Law and Economics' Equal Incompetence*, 91 GEO. L.J. 67, 73–74 (2002) (explaining that there are "individual and situational differences in rational behavior" and noting that one should not "assume [a] uniformity in cognitive performance across persons and situations that is not supported by empirical data"). Nonetheless, the unconscious tendency of most people to take these cognitive shortcuts is well documented in a number of situations applicable to jury decisionmaking, and understanding these tendencies and how to counteract them can vastly improve the way that experts present and juries understand information.

192. Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471, 1477–78 (1998).

193. *See, e.g.*, Christopher Jepson et al., *Inductive Reasoning: Competence or Skill*, 6 BEHAV. & BRAIN SCI. 494, 498 (1983) (discussing studies indicating that training in reasoning improves performance dramatically); Richard E. Nisbett et al., *Teaching Reasoning*, 238 SCI. 625, 630 (1987) (advocating formal training in the "rules underlying reasoning"). Training is much more effective for pragmatic applications of reasoning than for abstract principles. *See* Patricia W. Cheng et al., *Pragmatic Versus Syntactic Approaches to Training Deductive Reasoning*, *in* RULES FOR REASONING 165, 186 (Richard E. Nisbett ed., 1993) (noting that the "near total ineffectiveness of purely abstract training in logic contrasts dramatically with the ready ease with which people seem able to apply a naturally acquired pragmatic reasoning

that juries are not. Repeat exposure to the decisionmaking task, at least in the presence of feedback, can improve performance.[194] Third, judges are accountable, at least in some respects; they must present the basis for their decisions, and that basis is subject to scrutiny and to being overruled by a higher court. Accountability, which refers to the expectation that one may have to justify one's actions,[195] can improve judgment in the context of competing inferences from contradictory information.[196]

a. Processing Limitations

Although brain structure and function limit human memory and attention, people tend to make inferences as though both were infallible, resulting in cognitive shortcuts.[197] These tendencies are unconscious processes and doubtless enable people to make decisions that are fast and, on average,

---

schema" and noting that people "who received a brief training session on the obligation schema improved markedly on selection problems interpretable in terms of that schema").

194. *See, e.g.*, Jonathan J. Koehler, *The Base Rate Fallacy Reconsidered: Descriptive, Normative and Methodological Challenges*, 19 BEHAV. & BRAIN SCI. 1, 6 (1986) (citing studies showing that people learned to use base rates more effectively after receiving feedback about their errors from their experience (for example physicians who learned the low base rate of pneumonia from their practice experience relied heavily on the base rate when making diagnoses, and auditors "learned and used the base rate for financial statement errors most easily by directly experiencing those errors"), but cautioning that "personally experienced base rates were used only by those who also experienced the relationship between the base rate and the diagnostic information").

195. Jennifer S. Lerner & Philip E. Tetlock, *Accounting for the Effects of Accountability*, 125 PSYCHOL. BULL. 255, 255 (1999).

196. That is, accountability is effective only if the supervisor can access contrary information and disagree with the conclusions drawn. *Id.* at 258. These researchers note that accountability may have complex effects on judgment and that certain conditions must be present in order for accountability to have a positive effect, principally "an audience (a) whose views are unknown, (b) who is interested in accuracy, (c) who is interested in processes rather than specific outcomes, (d) who is reasonably well-informed, and (e) who has a legitimate reason for inquiring into the reason behind participants' judgments." *Id.* at 259. While the views of superior courts may be somewhat known, judges have a wider audience consisting of legal scholars and practitioners to respond to their decisions, even if the judges are not directly accountable to them; accuracy and the process of judicial inquiry into scientific validity is (or at least is supposed to be) the focus of the supervising courts, which hopefully are well informed and which have a legitimate reason for inquiring into the reason behind judicial admissibility decisions. Thus, accountability is a further reason for structuring the process so that judges assume gatekeeping responsibilities with respect to expert predictions of dangerousness.

197. *See* Jeffrey L. Rachlinski, *The "New" Law and Psychology: A Reply to Critics, Skeptics, and Cautious Supporters*, 85 CORNELL L. REV. 739, 750 (2000) (observing that "people make inferences based on attention and memory as if these processes are infallible, even though both are error-prone").

accurate enough.[198] It also keeps people from being paralyzed into inaction.[199] Although this ability may confer an evolutionary advantage to humans as a species, it does not lead to optimal decisionmaking in all situations.[200]

For example, rational choice theory predicts that people consider the statistical probability that an event will occur and update it with particularized specific information. In fact, however, people ignore base rates[201] and overestimate the correlation between what something appears to be and what it is, adhering to stereotypes.[202] In other words, they are likely to base decisions on the extent to which a particular event (or person) resembles a certain category

198. *See* Gerd Gigerenzer & Daniel G. Goldstein, *Reasoning the Fast and Frugal Way: Models of Bounded Rationality*, 103 PSYCHOL. REV. 650, 651–55 (1996) (designing and empirically testing algorithms of bounded rationality against statistically rational algorithms to solve real world problems of limited knowledge and finding that the statisficing algorithms scored the highest proportion of correct inferences in the shortest time).

199. *See, e.g.*, SHELLEY E. TAYLOR, POSITIVE ILLUSIONS: CREATIVE SELF-DECEPTION AND THE HEALTHY MIND 212-14 (1989) (noting that the only people who do not suffer from overoptimism bias about their chances for success in the future are the clinically depressed).

200. *See, e.g.*, Robyn M. Dawes, *Behavioral Decision Making and Judgment*, *in* 1 HANDBOOK OF SOCIAL PSYCHOL. 497, 497 (Daniel T. Gilbert et al. eds., 4th ed. 1998) (explaining that if actual decisionmaking violates a certain principle of rationality "*systematically* (not just as a result of unreliability or 'error'), this deviation is termed an *anomaly*—if the people who violate these principles simultaneously accept them as ones that they believe *should* govern their decision making").

201. The base rate is the frequency of a given subject in the population. For example, if a sample of 100 people consists of 70 lawyers and 30 engineers, the base rate of lawyers is 70%, and of engineers, 30%. Knowing only that information, if you were asked the occupation of any given person, you would be wise to answer "lawyer." Interestingly, most people do not. In a study that divided subjects into two groups, both of which were told that 100 people were either lawyers or engineers, one subject group was told there were 70 lawyers and 30 engineers, the other group that there were 30 lawyers and 70 engineers, and both groups were given thumbnail descriptions of the people written by psychologists, designed to be nondiagnostic with respect to occupation. Nisbett et al., *supra* note 193, at 626. In both groups, the subjects based their answers on stereotypes rather than population base rates. *Id.* *But see* Koehler, *supra* note 194, at 3 (arguing that it is not so much that base rates are ignored as that "subjects attach *relatively less weight* to base rate information than to descriptive, individuating information"). Regardless of whether subjects ignore base rates or give them insufficient weight, however, when the goal is accuracy in judgment, structuring a decision process to minimize errors would appear advantageous. Koehler suggests avoiding the base rate fallacy by explicitly structuring the task to sensitize decisionmakers to the base rate, presenting the information in relative frequentist terms, and giving the decisionmaker cues to base rate diagnosticity, as well as invoking heuristics that focus on the base rate. *Id.* at 5. That is precisely what this Article advocates in requiring judicial gatekeeping that limits expert prediction testimony to that which explicitly refers to population base rates.

202. *See* Einhorn & Hogarth, *supra* note 33, at 416 (giving the example of a waiter who gives better service to well-dressed patrons because of a belief that well-dressed patrons leave better tips, which the patron who leaves a good tip validates).

of events (or fits within their stereotypes of people).[203] People think by association and respond to patterns—even infants recognize shape patterns.[204] This characteristic of human thought undoubtedly speeds up the thinking process, but it has a downside. For example, in assessing the career of a person described as overbearing, aggressive, rude, and skilled at rhetorical argument, people will refer mentally to known stereotypes rather than population base rates.[205] Thus, in assessing the probability of future violent behavior for a particular defendant, jurors are unlikely to know, much less refer to, the probabilities of a recurring incident of violence given that one incident has already occurred.[206] Yet that information is crucial to their task.

---

203. *See* Kahneman & Tversky, *supra* note 30, at 431 (defining the representativeness heuristic as evaluating "the probability of an uncertain event, or a sample, by the degree to which it is: (i) similar in essential properties to its parent population; and (ii) reflects the salient features of the process by which it is generated" so that it matches a characteristic to a category and evaluates the probability in terms of the closeness of the match). The classic study on this bias was that of Meehl and Rosen, who documented the degree to which psychiatric diagnoses made in staff meetings ignored population base rates. Paul E. Meehl & Albert Rosen, *Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores*, 52 PSYCHOL. BULL. 194, 216 (1955). A simple example is that "a politician of erect bearing walking briskly to the podium is likely to be seen as strong and decisive; this is an example of judgment by representativeness." Daniel Kahneman & Amos Tversky, *On the Reality of Cognitive Illusions*, 103 PSYCHOL. REV. 582, 582 (1996).

204. *See, e.g.*, Dawes, *supra* note 200, at 534 (explaining the problems of representativeness and pseudodiagnosticity as probabilistic fallacies); Roger Lecuyer & Christine Cybula, *Categorization of Geometric Figures Composed of Three or Four Elements by 3-Month-Old Infants*, 19 CURRENT PSYCHOL. COGNITION 221, 221–44 (2000) (noting studies showing that infants recognize geometric patterns).

205. Amos Tversky and Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 163, 164 (Daniel Kahneman et al. eds., 1982). Kahneman/Tversky and Gigerenzer maintain an ongoing debate about whether this is a reasoning error or not, with the focus of the disagreement on interpretations of probability. *See, e.g.*, Gerd Gigerenzer, *The Bounded Rationality of Probabilistic Mental Models*, *in* RATIONALITY: PSYCHOLOGICAL AND PHILOSOPHICAL PERSPECTIVES 284, 291–97 (K.I. Manktelow & D.E. Over eds., 1993) (arguing that one cannot assign probabilities to unique events and therefore that there is no normative basis for assigning error to stereotyping, and suggesting that one may eliminate errors by asking questions in terms of frequencies rather than in terms of probabilities and increasing the use of random sampling); Kahneman & Tversky, *supra* note 203, at 582–83 (acknowledging that representation in terms of absolute frequencies improves accuracy, but citing studies to demonstrate that people nonetheless perceive correlations that do not exist and that "some significant judgmental biases are not readily corrected by the observation of natural frequencies"). For purposes of our discussion, it is enough to note that both sides agree that parties rarely present information to decisionmakers in a form—frequency—that is optimal for accuracy.

206. *See* Dawes, *supra* note 200, at 532 (explaining the fallacy of "considering the probability of the evidence given the hypothesis . . . without looking at . . . the prior odds").

A related problem is that people frequently overestimate the relevance of memorable incidents at the expense of statistical base rates.[207] It is not surprising that people make judgments on the basis of what they remember.[208] The problem arises, however, when people think that what they remember is representative of the population as a whole. For example, medical student and physician participants greatly overestimated the number of dramatic deaths in a study that asked participants to estimate the number of deaths caused by each of forty-two diseases.[209] Thus, the vivid recounting in the media of the statistically rare incidents of violent recidivism may well disproportionately influence jurors in their decision.[210]

In addition, because the jurors have recently heard in graphic detail how the defendant committed one atrociously violent act, they will likely believe that it represents the way that the defendant will behave in the future. The representativeness heuristic suggests that jurors do not refer to base rates (of which they are usually ignorant in any event) in their decisionmaking process. The anchoring heuristic suggests that because the jurors first learned about the defendant in the context of a graphically violent crime, they are likely to persist in thinking of the defendant as violent, even in the face of contrary evidence.[211] People frequently arrive at a decision that varies according to their starting point.[212] The availability heuristic, the tendency of people to

---

207. Tversky & Kahneman, *supra* note 205, at 164 (describing the availability heuristic).

208. For example, "a judgment of the prevalence of suicide in a community is likely to be mediated by the ease with which instances come to mind; this is an example of the availability heuristic." Kahneman & Tversky, *supra* note 203, at 582.

209. *See* Jay Christensen-Szalanski et al., *Effects of Expertise and Experience on Risk Judgments*, 68 J. APP. PSYCHOL. 278, 278 (1983) (identifying a study showing that experts and non-experts make differently biased risk judgments because of their differing exposure to risky events).

210. Interviews with capital jurors, for example, found that jurors overwhelmingly underestimated the amount of time that a defendant would have to serve before becoming eligible for parole, relying primarily on memory of vivid media accounts of violent recidivism. *See* William J. Bowers & Benjamin D. Steiner, *Death by Default: An Empirical Demonstration of False and Forced Choices in Capital Sentencing*, 77 TEX. L. REV. 605, 671–72 (1999) (discussing the interviews of 916 capital jurors in the Capital Juror Project).

211. Anchoring is the tendency for arbitrary starting points to influence decisions. Daniel Kahneman et al., *Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues*, *in* CHOICES, VALUES AND FRAMES 642, 665–66 (Daniel Kahneman & Amos Tversky eds., 2000) (describing the anchoring effect as a response that "is strongly biased toward any value, even if it is arbitrary, that the respondent is induced to consider as a candidate answer" and noting that the anchoring effect is one of "the most robust observations of the psychological literature").

212. For example, when asked to estimate percentages of United Nations countries that are African after being exposed to the result of a spin on a wheel of fortune—an obviously random

confuse the facility with which they can recall an event with its likelihood of recurrence,[213] suggests that the jurors will confuse vivid recollection of a horrendous crime that the defendant committed with future propensities for violence.

These problems are not insoluble. It is possible to help the jury make more rational (less biased) decisions. The future dangerousness determination is key to the jury's life or death determination.[214] Giving jurors plain alternatives to inform their sentencing judgment would help.[215] Jurors' release estimates strongly correlate with their final punishment vote.[216] The issue of danger is intertwined with the question of "danger when?"[217] Studies invariably report significant decreases in violent behavior as people age.[218]

Making the probabilistic nature of the evaluation explicit tends to reduce the frequency of these errors.[219] Moreover, presenting probabilistic evidence in frequentist terms makes it more comprehensible.[220] People are capable of sound reasoning if a party presents them the information correctly.[221] Thus,

---

and irrelevant value—people responded with marked differences according to the value spun on the wheel (the median answer of people for whom the value on the wheel was 10 was considerably lower than that for people exposed to a wheel value of 65). Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124, 1128 (1974). Cognitive psychologists explain this seemingly irrational behavior as the anchoring effect, in which people make decisions according to some (perhaps irrelevant) starting value.

213. *See id.* at 163 (describing the availability heuristic and the cognitive biases that may result).

214. *See* John H. Blume et al., *Future Dangerousness in Capital Cases: Always "At Issue"*, 86 CORNELL L. REV. 397, 398–99 (2001) (observing, on the basis of interviews with over one 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").

215. *See* Bowers & Steiner, *supra* note 210, at 609 (arguing that keeping jurors uninformed about sentencing alternatives skews the decision in favor of death).

216. *See id.* at 665 (noting that "jurors who estimate release in 20 or more years are consistently and substantially less likely to vote for death than those who thought release would come in 0-9 years or 10-19 years").

217. *See* John F. Edens et al., *Psychopathy and the Death Penalty: Can the Psychopathy Checklist-Revised Identify Offenders Who Represent "a Continuing Threat to Society?"*, 29 J. PSYCHIATRY & L. 433, 451 (2001) (contending that the real question should be how dangerous the defendant will be upon release in twenty years or within the confines of an institution).

218. *See id.* at 444-45 (reporting that violence decreases as age increases).

219. *See* Gerd Gigerenzer, *From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology*, 98 PSYCHOL. REV. 254, 267 (1991) (explaining the value of a structured reasoning process).

220. *See generally* GERD GIGERENZER, CALCULATED RISK (2002).

221. *See* Koehler, *supra* note 194, at 15 (citing studies demonstrating that presenting information in certain ways improves people's capability of sound probabilistic reasoning).

if experts were to testify in such a way as to make the probabilistic nature of their assessment task both explicit and frequentist, neither of which they currently do, they would aid the jury in its task of making an accurate decision.[222]

b. *Egocentrism and Cognitive Dissonance*

A well-documented bias of individuals is a tendency to overrate their abilities and their control over events, at least when the questions are difficult and the decisionmakers have no prior experience in making such decisions.[223] Across cultures, people appear to overestimate their ability to provide correct answers to questions.[224] This overconfidence may have unforeseen consequences in the context of capital sentencing. People tend to believe that their judgment is correct.[225] In addition, the social environment of the jury deliber-

---

222. *See generally* Beecher-Monas & Garcia-Rill, *supra* note 39 (arguing that this feature of certain actuarial instruments—providing an explicit probabilistic assessment—should improve jury decisionmaking).

223. *See* Gigerenzer, *supra* note 205, at 297-300 (noting the results of two decades of research showing that test participants were overconfident when judging the correctness of their answers to difficult general knowledge questions, but when directed to assess their correctness with reference to their prior experience in answering similar general knowledge tests, their overconfidence disappears). The problem for jurors is that they do not have any such reference points when it comes to assessing the defendant's dangerousness. The major determinant of overconfidence is the difficulty of the question. *See* Lyle A. Brenner et al., *Overconfidence in Probability and Frequency Judgments: A Critical Examination*, 65 ORG. BEHAV. & HUMAN DECISION PROC. 212, 213 (1996) (observing that "the major (though not the sole) determinant of overconfidence is the difficulty of the questions"). Researchers have observed overconfidence in a number of predictive tasks, including physicians' predictions of disease, economists' forecasts of recession, and players' predictions of their opponents' moves. *See id.* at 213 (citing studies demonstrating overconfidence in predictive tasks). However, as one researcher points out, we do not know if the kind of question domain makes a difference, or "whether there are simply some domains in which we tend to exaggerate the accuracy of our knowledge or judgment (not in others)." Robyn M. Dawes & Matthew Mulford, *The False Consensus Effect and Overconfidence: Flaws in Judgment or Flaws in How We Study Judgment?*, 65 ORG. BEHAV. & HUMAN DECISION PROCESSES 201, 210 (1996). Nonetheless, although we do not know whether the results in the general knowledge questions apply equally to the decision that the jurors make in capital sentencing, if our goal is to improve accuracy, we should implement ways of minimizing such effects.

224. *See* Mitchell, *supra* note 191, at 132 (citing studies showing that Asians, with the exception of Japanese and Singaporeans, are even more overconfident than Westerners).

225. *See* Hart Blanton et al., *Overconfidence as Dissonance Reduction*, 37 J. EXPT'L SOC. PSYCHOL. 373, 373 (2001) (citing studies asking people to evaluate their ability in solving laboratory problems and showing that "people think that they can solve problems that they cannot, think that they have made progress toward correct solutions when they have not, and think that they have drawn correct conclusions when they have not").

ations may increase overconfidence, given that researchers have demonstrated greater overconfidence in people acting within small social networks.[226] Researchers characterize these networks as having three to fifteen members (a characteristic that juries, normally twelve members, share); someone in a central, coordinating position (here, the judge); and weak contact with outsiders (judges typically tell juries not to discuss the case with anyone).

A number of different explanations account for overconfidence bias, and each of these accounts has implications for jury decisionmaking and emphasizes the importance of judicial screening of expert testimony for accuracy. For example, people may "confuse easily drawn inferences for easily remembered facts."[227] In addition, people might selectively focus on evidence that is consistent with their first impression and ignore inconsistent evidence.[228] Finally, people's overconfidence is a buffer against anxiety.[229]

Moreover, cognitive dissonance theory suggests that people tend to take further actions that justify and reinforce decisions that they have already made.[230] For example, gamblers and voters are more confident after they have placed their bets or votes than they were before.[231] The jury has already decided that the defendant is a very dangerous person when it found the defendant guilty, and the overconfidence bias tends to make them more confident in their action—a guilty verdict—than the facts would suggest. Any information that supports their decision is likely to have a disproportionate impact on their sentencing decision because of the related phenomenon of

---

226. *See* Joshua Klayman et al., *Overconfidence: It Depends on How, What, and Whom You Ask*, 79 ORG. BEHAV. & HUM. DECISION PROC. 216, 243 (1999) (finding an overall bias toward overconfidence, particularly in small social networks).

227. Blanton et al., *supra* note 225, at 374 (citing studies).

228. *See id.* (citing studies). *See generally* Eddie Harmon-Jones & Judson Mills, *An Introduction to Cognitive Dissonance Theory and an Overview of Current Perspectives on the Theory*, *in* COGNITIVE DISSONANCE: PROGRESS ON A PIVOTAL THEORY IN SOCIAL PSYCHOLOGY 3, 3–19 (Eddie Harmon-Jones & Judson Mills eds., 1999) (citing studies demonstrating that people selectively seek information that will decrease expected post-decision dissonance).

229. *See* Jeff Greenberg et al., *Why Do People Need Self-Esteem? Converging Evidence that Self-Esteem Serves an Anxiety-Buffering Function*, 63 J. PERSONALITY & SOC. PSYCHOL. 913, 913–21 (1992) (arguing that self-esteem sustains the illusion of control and diminishes anxiety).

230. *See generally* LEON FESTINGER, A THEORY OF COGNITIVE DISSONANCE (1957). Festinger's theory provoked a great deal of controversy, but the empirical basis for it appears to have survived the controversy. *See, e.g.*, Dawes, *supra* note 200, at 557–61 (detailing the controversy and concluding that "cognitive dissonance theory is resilient").

231. *See* Blanton et al., *supra* note 225, at 374 (arguing that "overconfidence reflects the motive to maintain a view of the self as a knowledgeable perceiver who makes sound judgments" and citing studies demonstrating this proposition).

cognitive dissonance.[232] This outcome appears to be what happens in capital sentencing: capital jurors overwhelmingly focus on the question of guilt even after the verdict when they are supposed to focus on the separate question of the appropriate punishment.[233] Overconfidence and cognitive dissonance suggest that jurors may place too much confidence in their decision of guilt and subsequently overvalue the expert prediction that confirms their decision, giving disproportionate weight to any information that confirms their initial decision of guilt.[234] This reason further justifies excluding testimony unsupported by data, such as when an expert testifies with more certainty than is warranted (i.e., the statement that he can be "one-hundred percent certain" that the defendant will kill again).

c. The Dilution Effect

The dilution effect is a cognitive bias that occurs when people receive complex information, some of which is relevant to the decision task and some of which is irrelevant.[235] When people are asked to judge whether someone else has a particular characteristic, such as aggressiveness, their judgments tend to be more focused (and accurate) when they receive only relevant information; even obviously irrelevant details appear to affect judgment.[236]

The dilution problem arises because irrelevant information obscures relevant information.[237] It may also arise because, in a testing context, people

---

232. Cognitive dissonance is a phenomenon in which people will adjust their attitudes and beliefs in order to justify a previously undertaken decision or course of action. *See generally* FESTINGER, *supra* note 230.

233. *See* Ursula Bentele & William J. Bowers, *How Jurors Decide on Death: Guilt Is Overwhelming; Aggravation Requires Death; and Mitigation Is No Excuse*, 66 BROOK. L. REV. 1011, 1017-19 (2001) (describing the capital juror project and its findings that jurors continue to focus on guilt after the verdict and that they appear to ignore, discredit, and devalue mitigating evidence even when it appears extensive and credible).

234. *See, e.g.*, Ziva Kunda, *The Case for Motivated Reasoning*, 8 PSYCHOL. BULL. 480, 480 (1990) (observing that people will often construct theories and use evidence in ways that make their final inferences come out the way that they want them to).

235. Some of the classic work on the dilution effect was that of Philip E. Tetlock and his co-authors. *See, e.g.*, Philip E. Tetlock et al., *The Dilution Effect: Judgment Bias, Conversational Convention, or a Bit of Both?*, 26 EUR. J. PSYCHOL. 915, 916-17 (1996) (citing studies demonstrating that "linking diagnostic with nondiagnostic evidence produced more regressive predictions than people would otherwise have made").

236. *See* Dawes, *supra* note 200, at 532, 537 ("*Dilution effects* occur when evidence that does not distinguish between hypotheses in fact influences people to change their mind.").

237. "Decisionmakers allow irrelevant information to alter the decisions that they would otherwise choose by considering only relevant information." Mark Seidenfeld, *Cognitive Loafing, Social Conformity, and Judicial Review of Agency Rulemaking*, 87 CORNELL L. REV. 486, 502 (2002) (using the example of a guess about a student's grade point average, in which

focus on social cues and assume that the experimenter would not present them with information unless he expected them to consider it.[238] Irrelevant information ought to be ignored, but it is not.[239] The dilution effect explains the harmful consequences of simply permitting the jury to hear expert testimony and then try to sort out its relevance for themselves.

d. The Dynamics of Group Decisionmaking

In addition to the dilution effect, the dynamics of group decisionmaking also bear on why the *Barefoot* court incorrectly believed that the jury deliberation is the right phase of the proceeding to sort out good science from bad. The jury's sentencing determination is unanimous, a group decision reached after deliberation; it is based on ideals of deliberative democracy that argument and reflection among competing views will lead to better—more accurate—outcomes.[240] At the same time, however, a single dissenter can shift the result to a hung jury. Thus, the process has aspects both of group consensus reaching and interactive individual decisionmaking.

Although both individual and group decisionmaking are subject to biases—decisions about what and how information is relevant[241]—some characteristics of group decisionmaking emphasize the importance of protect-

---

people gave very different predictions about a student's grade point average when told only the number of hours studied weekly than when also told how many plants the student kept). There are two major explanations for this effect, the representativeness heuristic, in which people select outcomes depending on the degree to which the evidence fits stereotypical patterns, and norms about social discourse, in which the experiment participants expect that information given them is relevant to their task. *Id.*

238. *See* Tetlock et al., *supra* note 235, at 916 (observing that "[f]ar from representing an error, the dilution effect may constitute a rational response to the interpersonal and institutional demands that impinge on individual perceivers").

239. *See* Dawes, *supra* note 200, at 537 (noting that "[s]omeone who has a strong opinion based on very strong evidence may be influenced to 'moderate' this opinion by exposure to a flurry of uninformative information").

240. *See* Cass R. Sunstein, *Deliberative Trouble?, Why Groups Go to Extremes*, 110 YALE L.J. 71, 73–75 (2000) (noting the received view that group deliberation yields an outcome that takes everyone's position into account and contending that, on the contrary, people deliberating in groups tend to polarize their positions from that of any individual member).

241. *See* Chip Heath & Rich Gonzalez, *Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence Against Information Collection Views of Interactive Decision Making*, 61 ORG. BEHAV. & HUM. DECISION PROC. 305, 323 (1995) (noting that individual interactive decisionmaking exhibits similar characteristics to group consensus decisionmaking); Norbert L. Kerr et al., *Bias in Judgment: Comparing Individuals and Groups*, 103 PSYCHOL. REV. 687, 714–15 (1996) (defining bias as reflecting "decisions about whether and how to use information" and demonstrating that "groups will amplify bias under some conditions but attenuate it under others").

ing the jury from irrelevant information that exacerbates biases and providing instructions that can guide the group's reasoning process. Group decisions are better than individual decisions when evaluating information that has a demonstrably correct solution.[242] The reason is that the errors of individuals in assessing information tend to cancel each other out.[243] Thus, I do not argue that the judicial system should replace the jury in criminal trials or in capital sentencing proceedings. I do suggest that the confluence of systematic errors in the context of sentencing proceedings requires that judges carefully screen information that the jury will use to make its collective decision.

The argument that collective decisionmaking should cancel out judgmental errors does not work for systematic biases.[244] Rather, collective processes under certain conditions skew the decision away from judgmental accuracy.[245] These conditions are precisely those that a jury faces in making its determination—for example, predictive judgment tasks, tasks in which jurors have no clearly shared framework for defining right or wrong answers.[246]

One anomalous tendency of group decisions is that groups often polarize; that is, the group will make a more extreme decision than the initial position of any individual in the group would have predicted.[247] The result is that

---

242. *See* Daniel Gigone & Reid Hastie, *Proper Analysis of the Accuracy of Group Judgments*, 121 PSYCHOL. BULL. 149, 149 (1997) (noting that when one group member can justify correct answers, the rest of group usually decides the correct answer).

243. *See id.* at 159 (observing the cancellation of uncorrelated errors in group decisionmaking).

244. *See* Kerr et al., *supra* note 241, at 713–14 (noting that although the law of large numbers suggests that random errors will cancel each other out in collective decisions, it will not do so for systematic errors).

245. *See id.* at 714–15 (citing studies demonstrating the benefits of group over individual decisionmaking).

246. *See, e.g.*, Garold Stasser et al., *The Social Psychology of Jury Deliberations: Structure, Process and Product*, *in* THE PSYCHOLOGY OF THE COURTROOM 221, 221–56 (Norbert L. Kerr & Robert M. Bray eds., 1982) (providing studies demonstrating various factors' effects upon the jury decisionmaking process).

247. *See, e.g.*, Daniel J. Isenberg, *Group Polarization: A Critical Review and Meta-Analysis*, 50 J. PERSONALITY & SOC. PSYCHOL. 1141, 1141 (1986) (noting that "an initial tendency of individual group members toward a given direction is enhanced following group discussion"). A number of theories explain this tendency, such as social comparison theory (people initially espouse opinions less extreme than their true opinion because they fear being labeled deviant, and once they realize that others have more extreme opinions they shift theirs to the true value); persuasive arguments theory (explaining group polarization on the basis of a pool of arguments drawn from discussion among the group members); self-categorization theory (group members define the social identity of the group and then modify their positions to conform with it); social influence network theory (a network of interpersonal influence); and social decisions schemes (the distribution of initial opinions specifies the relative influence of the alternative initial positions of group members). *See* Noah E. Friedkin, *Choice Shift and*

group decisions tend to coalesce around an extreme position rather than around the middle of antecedent positions.[248] In mock-jury studies, researchers have extensively documented polarization around the question of guilt or innocence.[249] In interactive groups, group members tend not to respond to information against their position by modifying their position or lowering their confidence. Instead, researchers have found that interaction increases group members' confidence in their decision in a way that is unjustified by increased accuracy.[250] Group members frequently fail to respond to the information presented.[251]

The term "social loafing" describes the observed phenomenon that individuals put less effort into a group decision than into an individual decision.[252] The classic experiments deal with physical exertion, such as rope pulling, shouting, and clapping, and at all these tasks, group output tends to

---

Group Polarization, 64 AM. SOC. REV. 856, 857–60 (1999) (discussing alternative explanations). Discussion of the alternative explanations is well beyond the scope of this Article.

248. *See* Friedkin, *supra* note 247, at 857 (explaining the concept of group polarization in terms of a choice shift, which occurs "when, after a group's interaction on an issue, the mean final opinion of group members differs from the members' mean initial opinion . . . in the opposite direction of the initial inclination of the group").

249. *See* ROGER BROWN, SOCIAL PSYCHOLOGY 227-29 (2d ed. 1986) (collecting studies). Very diverse groups tend to diminish this effect. Sunstein, *supra* note 240, at 116 (citing James Fishkin's experiments demonstrating an absence of polarization effects in groups composed of highly diverse individuals). In one respect, however, capital juries are not heterogeneous at all: in order to serve on the jury, jurors must be willing to support the death penalty (in the appropriate case, as the voir dire usually instructs juries). Counter-intuitively, it might be less polarizing to have a diversity of opinions about the legitimacy of the death penalty.

250. *See* Heath & Gonzalez, *supra* note 241, at 306 (arguing that interaction does not cause people to assess the available information differently but merely to develop more coherent rationales for their choices and beliefs). Heath and Gonzalez studied interactive decisionmaking—individual decisions made after consultation with the group—and distinguished it from group decisionmaking on the basis that groups must reach a consensus and the "aggregation procedure may hide or distort changes in individual preferences." *Id.* at 307. Jury decisionmaking has facets of both interactive and group decisionmaking; although the end product must be a consensual decision, in order to avoid a hung jury, each juror must individually agree, and each can hold out until persuaded. Moreover, Heath and Gonzalez concluded that consensus decisionmaking is not the only kind of group decision that exhibits the characteristics of groupthink. *Id.* at 323. The characteristics of groupthink, "'discount[ing] warnings and other forms of negative feedback that, taken seriously, might lead the group members to reconsider their assumptions,' . . . provide a remarkably satisfying description of the phenomenon . . . of individual decision makers interacting in a social environment." *Id.*

251. *See id.* at 323 (finding it "much harder to find evidence of information collection" in group studies conducted, a difficulty that surprised the authors).

252. *See* Seidenfeld, *supra* note 237, at 511 ("Social loafing occurs when an individual participates in producing a group product; such loafing has been established in the context of group decisionmaking.").

be less than the sum of the individuals' efforts when performing the same task alone.[253] Researchers have extended this work to judgment tasks, with similar results.[254] The reasons given for social loafing have to do with lack of accountability, inability to measure individual input, and lack of control over the output.[255] When, for example, researchers told test participants that they would have to justify their judgments, they exerted as much cognitive effort as individual judges.[256]

Moreover, pressures for uniformity and group loyalty can build up in collective decision processes to the point at which they adversely affect both cognitive efficiency and moral judgment.[257] Irving Janis analyzed a number of political decisions and concluded that group dynamics could lead to policies that "*deserved* to be fiascoes."[258] Janis found that group members in these "fiascoes" were so motivated to maintain each other's respect that it inhibited critical thought and dissenting opinions.[259] Groups isolated from outside influences and lacking systematic procedures for evaluating evidence were especially prone to this kind of cognitive error.[260] Stressful conditions further magnified these effects.[261] As a result, in the fiasco situations, decisionmakers

253. *See* Bibb Latane et al., *Many Hands Make Light the Work: The Causes and Consequences of Social Loafing*, 37 J. PERSONALITY & SOC. PSYCHOL. 822, 822–23 (1979) (discussing the Ringelman effect under these testing conditions). Although scholars have documented social loafing under diverse conditions, it appears to disappear under conditions of accountability. For example, social loafing is virtually eliminated when people learn that their individual production will be measured, even when working in groups. Kenneth L. Bettenhausen, *Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed*, 17 J. MGMT. 345, 361 (1991).

254. *See* Bettenhausen, *supra* note 253, at 361 (citing studies showing that people who shared responsibility for a complex judgment task put less cognitive effort into the task than individuals working alone).

255. *See id.* at 360–61 (explaining that "the phenomenon appears to center on the identifiability of the actor . . . . [S]ocial loafing was virtually eliminated when undergraduates were told their individual noise production would be measured . . . . [and] the potential for evaluation . . . increased motivation whether subjects could self-evaluate or not.").

256. *See generally* Kenneth H. Price, *Decision Responsibility, Task Responsibility, Identifiability, and Social Loafing*, 40 ORG. BEHAV. & HUMAN DECISION PROC. 330–45 (1987).

257. *See* Philip E. Tetlock et al., *Assessing Political Group Dynamics: A Test of the Groupthink Model*, 63 J. PERSONALITY & SOCIAL PSYCHOL. 403, 403 (1992) (performing multiple regression analysis of various collective decision processes and finding ample support for the groupthink hypothesis first advanced by Irving Janis).

258. IRVING L. JANIS, GROUPTHINK 9 (2d ed. 1982).

259. *Id.*

260. *See id.* at 199 (listing insulation of policy-making groups as one of the three conditions contributing to groupthink).

261. *See id.* at 258–59 (finding higher propensity for groupthink under stressful condi-

began to experience excessive optimism about the correctness of their judgment, silencing deviant opinions in order to re-establish consensus. The kinds of cognitive errors from this kind of group pressure include: (a) truncated consideration of alternatives and objectives (often, the group discussed only the option initially favored by group members), (b) a failure to examine the risks of the initially preferred choice, (c) a failure to reappraise initially rejected alternatives, (d) poor search for relevant information, (e) biased processing of information, and (f) a failure to work out contingency plans in the event that known risks materialized.[262]

The situation for group decisionmaking is not irremediable, however. Janis contrasted the "groupthink" conditions with conditions (he called these vigilance conditions) that led to more rational outcomes. Critical appraisal and open discussion of options made for more careful analysis. Group final judgment depends on both the relative beginning positions of the individual members and the processes by which the group combines preferences to define a group decision.[263] The importance of screening expert testimony for relevance and reliability has to do with focusing the starting point of deliberations. If the individual members recognize the normative use of particular information, the fact that the group is more likely to use that information properly emphasizes the importance of clear jury instructions.[264]

D. Proposed Solution: Gatekeeping in Capital Sentencing Proceedings

Judicial gatekeeping—screening for accuracy before permitting expert testimony—makes for more accurate judgments for a number of reasons.[265] As Robert Burns explains, the proper question with regard to admissibility is whether the evidence at issue would throw the jury "off track" in its goal of

---

tions).

262. Tetlock et al., *supra* note 257, at 404 (citing Janis).

263. *See* Kerr et al., *supra* note 241, at 694 (drawing mathematical conclusions in group decisionmaking).

264. *See id.* at 715 (arguing that guidance in principles of rational judgment improves collective decisionmaking).

265. As discussed earlier in this Article, to call something an "accurate" judgment is a normative statement that raises complex issues about what we know, how we know it, and what our goals are. In the context of the goals of a capitol juror, the jury appears to focus on the goal of determining whether this defendant would kill again if released. Interviews with capital jurors reflect this overwhelming concern. *See* Bowers & Steiner, *supra* note 210, at 665–67 (discussing the Capital Juror Project). The arguments of lawyers and legal scholars that the question ought to be whether the defendant will pose a threat to other inmates or prison personnel are beside the point. My point is simply that in light of their goal, the presentation of irrelevant (unscientific) expert testimony makes their determination less accurate.

reaching the public truth.[266] The dilution effect explains the importance of screening irrelevant information from jury decisionmaking. People participating in experiments expect that information given to them for the purpose of making a decision is relevant.[267] Similarly, jurors, who already participated in the guilt phase of the trial, have experienced judicial screening and may expect that whatever information they receive is relevant to their task.[268] When it is not relevant, the jury's decision will be less accurate than it would be if the court had excluded such evidence.

The dilution effect disappears, however, when people are made accountable for their judgment—as judges are by judicial review—as long as the decisionmaker knows that parties are presenting both relevant and irrelevant information.[269] In contrast, the dilution effect occurs even when one tells unaccountable individuals (jurors are unaccountable because they never need to explain the reason for their decision[270]) that they are receiving both relevant and irrelevant information and that they need to sort through it to reach their decision.[271] Irrelevant and inaccurate information throws group decisions off track even more than it does individual judgment. Thus, the dilution effect suggests that merely presenting evidence that counters the misinformation may not be enough.

Moreover, judges are experts because they make repeat decisions in the presence of feedback. People who repeat the decisionmaking process many times in the presence of feedback regarding their accuracy are more likely to

---

266. *See* ROBERT BURNS, A THEORY OF THE TRIAL 22 (2001) (discussing the need for materiality in evidence).

267. *See* Tetlock et al., *supra* note 235, at 915 (studying dilution effect).

268. For example, Tetlock and his co-authors found that test participants, even when told that a computer randomly generated the information and that it contained both relevant and irrelevant information, still made more regressive estimates given the irrelevant information. *See id.* at 926–27 (discussing research results).

269. *See id.* at 930-31 (demonstrating that "the dilution effect disappears among accountable subjects who were explicitly told that conversational norms did not apply because the information they had been given had been randomly selected from a computer database . . . [or] conversational norms were explicitly deactivated").

270. Even if one assumed that jurors were accountable, in the sense that they must render a public decision and may have to explain their reasons to friends and family, it is only the accountability to unknown audiences that appears to affect the care with which they scrutinize information. *See* Philip E. Tetlock & Richard Boettger, *Accountability: A Social Magnifier of the Dilution Effect*, 57 J. PERSONALITY & SOC. PSYCHOL. 388, 388 (1989) (defining accountability).

271. *See* Tetlock et al., *supra* note 235, at 931 (demonstrating that "explicitly deactivating conversational norms was not sufficient to eliminate the dilution effect among unaccountable subjects").

make accurate judgments.[272] Judges get more feedback than juries through the appellate process and through legal scholarship and commentary. Expertise tends to decrease both technical errors and the consideration of irrelevant information.[273] Experts (such as judges) trained in decision rules (such as the analysis required under *Daubert*) tend to make better judgments about validity than lay people (jurors) who are unaware of these rules.[274] Training can improve reasoning.[275] Judges, who have extensive training in legal analysis and, post-*Daubert*, in reasoning about expert testimony, can make better evaluations of such testimony than untrained jurors. In addition, although the overconfidence bias may afflict experts more than novices,[276] groups are more prone to it than individuals.[277]

Further, accountability can significantly improve the quality of some kinds of judgment.[278] Individuals making judgments under conditions of accountability are more likely to be careful and thoughtful than groups making judgments without individual accountability.[279] Judges are accountable not only to their superior courts but also to a wider audience of legal scholars and practitioners who will comment on their decisions. Accountability to an

---

272. *See* Rachlinski, *supra* note 197, at 756 (citing feedback as a prerequisite to learning).

273. *See* Seidenfeld, *supra* note 237, at 499 (discussing biases mitigated by expertise).

274. *See* Lerner & Tetlock, *supra* note 195, at 263 (explaining that accountability, which may attenuate biases resulting from lack of effort or self-critical awareness, has no effect on judgment tasks requiring knowledge of formal decision rules that are unfamiliar to the decisionmaker).

275. *See* Mitchell, *supra* note 191, at 87 n.48 (citing studies showing the extent to which training increases the accuracy of decisionmaking).

276. *See* Seidenfeld, *supra* note 237, at 498 (noting that overconfidence in predictions is a "bias to which experts may be more prone than novices").

277. *See* Tetlock et al., *supra* note 257, at 418–19 (noting that "groupthink" promoted rigid and self-righteous patterns of thinking).

278. *See* Shelley E. Taylor, *The Social Being in Social Psychology*, *in* 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY 58, 76 (Daniel T. Gilbert et al. eds., 4th ed. 1998) ("Accountability for one's inferences produces more thorough and more elaborate processing that takes account of more information and that is, at least sometimes, more accurate than processing that occurs in the absence of accountability."). Because the conditions of judicial gatekeeping review are precisely those that increase accuracy, it is the judge rather then the jury who should be the locus of the decision about admissibility of expert testimony. *Cf.* Seidenfeld, *supra* note 237, at 509–10 (explaining that judicial review of agency decisionmaking falls within the definition of accountability because courts examine the arguments pertaining to the validity of agency reasoning). Like judicial review of agency decisions, judicial review of gatekeeping determinations similarly examines the basis on which the judge made the determination.

279. *See* Bettenhausen, *supra* note 253, at 361 (citing studies demonstrating that people in groups of sixteen who shared responsibility for the judgment task "used less complex judgment strategies than subjects working alone," although "multiple judges who expected to justify their judgments worked as hard as individual judges").

unknown audience enhances careful decisionmaking.[280] People who know that they will have to justify their decisions ahead of time, as judges do, perform better cognitively.[281] Moreover, when the accountability review evaluates the process resulting in the judgment rather than the outcome, those process judgments improve under conditions of accountability.[282] The gatekeeping decision of the judge is reviewable primarily for its process rather than its outcome,[283] so one would expect a more careful and critical evaluation of the evidence than one could expect from jurors, who at most may have some outcome accountability in terms of possible negative consequences from their community.

In addition, some cognitive reasons, apart from the empirical invalidity of clinical judgment, show that actuarial testimony presented as a frequency determination is apt to be less skewed than a subjective accuracy determination. People making decisions, whether jurors or judges, make better decisions when presented with frequentist probabilities rather than subjective probabilities.[284] People, for example, asked how many times out of ten they are likely to choose the wrong answer are more likely to be accurate, even when they assess their accuracy at ninety-five percent.[285] Thus, not only is actuarial testimony better science, it is better for the decisionmaking process.

Gatekeeping is an important first step, but clear jury instructions are also a factor. Studies demonstrate that jurors lack understanding of how to weigh the evidence presented in death penalty proceedings.[286] Jurors simply do not

---

280. *See* Lerner & Tetlock, *supra* note 195, at 256 (explaining that "people often seek approval from their respective audience," and if "audience views are known prior to forming one's own opinion, conformity becomes the likely coping strategy," and the result is likely to decrease rather than increase accuracy as it does when the person making the opinion does not know the audience).

281. *See id.* at 257 (acknowledging that postdecisional accountability leads to self-justification rather than self-criticism and thus poor decision performance, but stating that people who know that they will be held accountable before engaging in the judgment task tend to be highly self-critical and more accurate).

282. *See id.* at 258 (citing studies showing that "accountability for decision outcomes—rather than decision processes—would increase the escalation of commitment to prior courses of action . . . [while] [p]rocess accountability, by contrast, would (a) lead decision makers to engage in more evenhanded evaluation of alternatives and (b) decrease the need for self-justification").

283. *See* Daubert v. Merrell Dow Pharms., 509 U.S. 579, 589 (1993) (discussing the admissibility of expert testimony).

284. *See* GIGERENZER, *supra* note 220, at 7 and *passim* (urging communication of risk in frequentist terms).

285. *Id.*

286. *See* James Luginbuhl & Julie Howe, *Discretion in Capital Sentencing Instructions: Guided or Misguided?*, 70 IND. L.J. 1161, 1176–77 (1995) (demonstrating how confusing jury

know how to assess the expert testimony.[287] These misunderstandings play a pivotal role because jurors use juror instructions as persuasive devices.[288] Frequently, jurors think that the instructions require them to impose death unless they can conceive of a reason not to do so.[289] Thus, not only do judges need to exercise their gatekeeping powers, they need to clarify the instructions that they give the jury.

VI. Gatekeeping Risk: Scientific Validity of Actuarial Testimony

A. Complexity Theory and the Inherent Limits of Prediction

The precise prediction of future behavior is impossible.[290] At best, predictions in complex systems are highly contingent, and human behavior is a paradigmatic complex system.[291] Complexity theory explains that human individuals are interacting parts of a complex world, interacting with our environment and other creatures, and that each brain originating behavior is itself a complex organ.[292] The most successful predictions are of weather phenomena, and even there, small changes in underlying conditions can have a huge impact on the result.[293] This impact is because small errors in determin-

---

instructions systematically predispose jurors toward a sentence of death).

287. *See* Joseph L. Hoffmann, *Where's the Buck?—Juror Misperception of Sentencing Responsibility in Death Penalty Cases*, 70 IND. L. J. 1137, 1149–50 (1995) (citing Capital Juror Project interviews of jurors who found death penalty instructions confusing, particularly with regard to how they were supposed to evaluate aggravating and mitigating factors).

288. *See* Luginbuhl & Howe, *supra* note 286, at 1177 (analyzing the effect of ambiguous jury instructions in capital cases).

289. *See* Hoffman, *supra* note 287, at 1152 (describing the statements of capital jurors).

290. *See* Paul E. Plsek & Trisha Greenhalgh, *The Challenge of Complexity in Health Care*, 323 BRIT. J. MED. 625, 625 (2001) (explaining the problem of prediction in terms of complexity theory).

291. "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." *Id.* (giving as examples the immune system, a colony of termites, the financial market, and "just about any collection of humans").

292. *See* ILYA PRIGOGINE, THE END OF CERTAINTY: TIME, CHAOS AND THE NEW LAWS OF NATURE 4-5 (1996) (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"); Mark D. Albertson, *Can Violence Be Predicted? Future Dangerousness: The Testimony of Experts in Capital Cases*, 3 CRIM. JUST. 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").

293. We cannot predict the weather because it is a "classic case of chaotic behavior." RICHARD SOLE & BRIAN GOODWIN, SIGNS OF LIFE 9 (2000). The impossibility of prediction is

ing the initial conditions (i.e., the predictors) may yield large errors in calculating expected outcomes, and even when one understands the properties of the individual components, the behavior of a system with many interacting components is inherently unpredictable.[294] Weather patterns display both order and chaos, so 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.[295] This analysis has implications for predicting human behavior, especially over a lifetime.

Measuring initial conditions is crucial in making predictions about complex systems such as the weather and human behavior. In weather prediction, for example, "over 10,000 land-based stations and hundreds of ships collect weather information daily at six-hour intervals."[296] Nothing monitors human behavior in such detail. We often do not know the relevant factors about a defendant's environment (internal or external), and even what factors are relevant is hotly debated.

Moreover, making accurate predictions requires expertise—that is, accountability, feedback, and opportunities for repeat performance. Weather experts, for example, make their predictions based on information from weather stations, satellites, balloons, aircraft, and human spotters making daily observations, all of which information funnels to one of several meteorological centers.[297] These centers generate regional reports, which are then adapted to local conditions.[298] This process means that one can generate and analyze a huge amount of information at frequent intervals, something one unlikely can achieve in human behavior.

Further, even using all available data, there are limits to prediction.[299] There are three reasons for these limits. First, the human brain is the premier

---

because "small errors in initial conditions give rise to very large errors in calculating expected outcomes." *Id.* at 12. Complexity theory, the study of nonlinear systems (like weather), involves both the study of chaos, with sensitivity to initial conditions that makes dynamics unpredictable, and emergent properties, in which the general inability of observers to predict the behavior of nonlinear systems from their parts and interactions. *See id.* at 20 (analyzing parallels between chaos and emergent properties).

294. *See id.* at 13 (discussing scientific theory of reductionism).

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

296. John Monahan & Henry J. Steadman, *Violent Storms and Violent People: How Meteorology Can Inform Risk Communication in Mental Health Law*, 51 AM. PSYCHOL. 931, 933 (1996).

297. *See id.* (explaining the collection of data by weather services).

298. *Id.*

299. *See* Robinson, *supra* note 172, at 1450 (describing the inability of science to predict criminality even with all available data).

example of nonlinearity—that is, there is no predictable relationship between cause and effect.[300] The brain 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.[301] Second, an individual's conduct results both from internal stimuli and from stimuli from the environment, including a web of relationships affecting beliefs, expectations, and behavior.[302] 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 one can only consider probabilistic. Seemingly inconsequential stimuli can radically alter the equilibrium of each of these systems.

B. Actuarial Instruments: Admissible Under a Best Evidence Theory?

Because of the inherent limits of predictability in complex systems, the most that can be said is that actuarial instruments may improve on the woeful

---

300. *See* SOLE & GOODWIN, *supra* note 293, at 1 (distinguishing the linear relationship between cause and effect from nonlinearity).

301. 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 behavior 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 behavior. 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 to 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, 685 (2001) (citations omitted).

302. *See generally* ANTONIO R. DAMASIO, DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN (2000) (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).

inadequacy of clinical predictions.[303] Is that enough to get them through the gate of scientific validity analysis? My answer is a tentative yes, based on the Popperian notion that what makes a theory scientifically valid is its explanatory power.[304] Explanatory power is the ability of a theory, model, or hypothesis to take into account all the observed data and to make a persuasive scientific argument.[305] Unlike clinical diagnostic predictions, the explanatory power of actuarial instruments rests on the idea of risk analysis, a statistical methodology commonly used by epidemiologists, toxicologists, the Environmental Protection Agency, and the insurance industry, among others.

The theory underlying actuarial instruments is that structured reasoning processes improve accuracy of judgment. Actuarial instruments do not abandon human judgment; they simply structure it into a formal reasoning process.[306] Although some evidence suggests that a multidisciplinary team may be able to rival the accuracy of actuarial instruments, empirical data demonstrates that such structured analysis improves decisionmaking considerably.[307] Using structured analysis offers many advantages in human decisionmaking, particularly in light of the difficulty that people have in synthesizing differently weighted likelihoods of varying significance, such as risk factors for violent behavior.[308] Thus, actuarial instruments may offer a distinct advantage in assessing risk of violent behavior.[309]

---

303. *See* WEBSTER ET AL., *supra* note 35, at 20–21 (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").

304. *See* Beecher-Monas, *supra* note 5, at 1588-89 (discussing the concept of explanatory power).

305. *See* 1 IMRE LAKATOS, THE METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES 142 (1978) ("A hypothesis, however novel in its intuitive aspects, will not be allowed to be proposed, unless it has novel empirical content in excess of its predecessor.").

306. *See* QUINSEY ET AL., *supra* note 31, at 65 (discussing the value of human judgment to actuarial prediction instruments); Dolan & Doyle, *supra* note 40, at 304 (observing that "[s]tructured clinical judgment represents a composite of empirical knowledge and clinical/professional expertise").

307. *See* Fuller & Cowan, *supra* note 39, at 286 (comparing multidiscipinary clinical judgment with actuarial approaches).

308. *See* EISER & VAN DER PLIGT, *supra* note 42, at 100 (observing that human decision "accuracy declines considerably when the number of features or the number or alternatives increases . . . . [R]eliability with which choice rules are used tends to decrease as the decision-maker's information load increases.").

309. *See* Norval Morris & Marc Miller, *Predictions of Dangerousness, in* CRIME AND JUSTICE: AN ANNUAL REVIEW OF RESEARCH 1, 1 (Michael Tonry & Norval Morris eds., 1985) (concluding that an actuarial approach is best suited to predict violent behavior).

Risk analysis is based on statistical concepts of correlation. Statistical analysis provides an important tool for examining whether theories correspond with observation.[310] One should not, however, permit statistical analyses to generate hypotheses about causation. One cannot understand statistical measurements within the context of the system under study.[311] Here, the context is human behavior, a quintessentially complex phenomenon.[312] And although actuarial instruments measure observed statistical correlations of violent behavior with factors such as past patterns of violence, age, and ability to form lasting relationships, very little links these factors to a theory of human violence. Moreover, one cannot expect statistics to provide an answer about any particular individual.[313] The most that one can say from even the best statistical analyses is that someone falls within a group that has a certain statistical propensity for violence.[314]

Further, actuarial instruments are risk assessment tools that combine a number of risk factors to achieve an overall "score" that ranks levels of risk.[315] Although researchers have analyzed various risk factors, and numerous studies

---

310. *See* Kenneth J. Rothman, Editorial, *Significance Questing*, 105 ANNALS INTERNAL MED. 445, 445 (1986) (discussing the value and limitations of statistical procedure to evaluate measurement error regarding observations).

311. *See* THEODORE COLTON, STATISTICS IN MEDICINE 117, 304 (1974) (explaining that a result, although it may be statistically significant, may still be medically meaningless, and cautioning against permitting statistical analyses to generate hypotheses).

312. Complexity theory recognizes the continuous interaction of individuals with the whole system of which they are a part and that they respond to both random events and dynamic phenomena. SOLE & GOODWIN, *supra* note 293, at 20 (noting the "continuous conversation between parts and wholes").

313. 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 & James M. Robins, *Epidemiology, Justice, and the Probability of Causation*, 40 JURIMETRICS J. 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 one can say is that exposure caused a certain statistical increase of disease over background levels.

314. The concept of probability to which I am referring is the idea of long-run relative frequency. *See* COLTON, *supra* note 311, at 63 (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." MARTIN STERNSTEIN, STATISTICS 15 (1994). "Virtually identical conditions" are hard to come by in observational studies of humans (such as the studies underlying the violence risk assessment tools). This difficulty is a pervasive problem for human studies, but one that does not necessarily undermine their validity. *See* Beecher-Monas, *supra* note 5, at 1604–07 (discussing scientific validity of human studies).

315. *See* Monahan, *supra* note 187, at 903 (evaluating risk assessment instruments).

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; one must analyze them in concert.[316] One of the significant problems with clinical judgment is an apparent inability to adjust predictions according to these interrelationships.[317]

Repeated studies of actuarial methods have demonstrated them to be superior to clinical judgment standing alone.[318] 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.[319] Sometimes the factors are not independent; an example would be anger and the inability to sustain relationships.[320] Moreover, in assessing risk one must consider the time period that the risk assessment covers, the circumstances for its implementation (prison for life, in the case of capital sentencing, and perhaps eventual release into the community after serving a minimum term of the life sentence[321]), and the individual's motivation to refrain from violence (including the motivation to comply with treatment).[322] Yet rarely do courts or actuarial instruments address these considerations.

A further complication is that rare events—and recurring violence is statistically rare, even among violent offenders—are inherently difficult to predict.[323] The most that one can say for any actuarial risk assessment instrument is that it can give a probabilistic estimate of the level of risk for people who share characteristics with the person assessed.[324] And the estimate is subject to a great deal of error.[325]

---

316. *See* John Monahan, *Clinical and Actuarial Predictions of Violence: Scientific Status, in* 1 THE LAW AND SCIENCE OF EXPERT TESTIMONY § 9-2.1.1, at 425 (2d ed. 2002) (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").

317. QUINSEY ET AL., *supra* note 31, at 56.

318. *See* Cunningham & Reidy, *supra* note 40, at 28 (recognizing that actuarial methods are superior to clinical methods in predicting human behavior).

319. *See* Carson, *supra* note 42, 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).

320. *Id.*

321. This result occurs if permitted by statute; in some states, life term is without parole.

322. Carson, *supra* note 42, at 261.

323. *See* WEBSTER ET AL., *supra* note 35, at 5 (noting that "rare events are always harder to forecast than frequent occurrences").

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

325. As the VRAG authors explain, "there is a predictive sound barrier" of some unknown dimension. QUINSEY ET AL., *supra* note 31, at 168. Although they assert that it is greater than

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.[326] Indeed, all science is value-laden, and risk assessment is not different in that regard.[327] A risk appraisal can inform, but cannot answer the ultimate question of whether a jury ought to sentence a particular individual to death or to life in prison.[328] It is something that the jury must weigh, assessing 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 mistakenly predicted to be violent.

Finally, a caveat. Even if actuarial testimony is demonstrably more accurate than clinical testimony, not all actuarial testimony will meet the standards of scientific validity. First, not all actuarial instruments have the same empirical foundation.[329] Second, experts may conclude more from the instruments than they warrant and testify outside the scope of valid inferences. This possibility is a particular problem when experts use a categorical rather than a statistical articulation of their results: dangerous/not dangerous rather than "this defendant falls within a group that has a __ probability of recurring violence." Third, when experts rely on factors in addition to the actuarial instruments, each of those factors must have a demonstrable empirical basis. For example, in *Barnette*, not only did the expert testify that he relied on the PCL-R for his opinion that the defendant posed a future danger to society, but he also testified that psychopaths are like "fake fruit" in that they may look normal but they are not, a statement wholly unsupported by any scientific studies.[330]

40% accuracy, they do not have the data to substantiate the assertion. *Id.*

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

327. *See* Beecher-Monas, *supra* note 5, at 1575–77 (noting that the scientific method and scientific conclusions are not completely objective).

328. *See* QUINSEY ET AL., *supra* note 31, 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).

329. *See id.* at 141–42 (noting that actuarial instruments use different variables); *see also* Erica Beecher-Monas & Edgar Garcia-Rill, *Danger at the Edge of Chaos: Predicting Violence in a Post-Daubert World*, 24 CARDOZO L. REV. __ (forthcoming 2003) (discussing three prevalent actuarial instruments and concluding that the most sophisticated by far is the VRAG).

330. Edens et al., *supra* note 217 (citing and quoting United States v. Barnette, Transcript, Sentencing Phase, File No. 3:97CR23-P, Feb. 5, 1998). The expert also testified that he based his opinion on the defendant's callousness, as demonstrated by the defendant's eating lunch during his discussion with the expert, and his inability to define compassion on an intelligence test, both similarly unscientific bases. *Id.*

C. *Assisting the Jury*

Gatekeeping standards require that even scientifically valid expert testimony be capable of assisting the jury in its deliberations.[331] 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.[332] Despite problems with the accuracy of violence predictions, actuarial predictions provide the best information available, and courts should admit them to help guide decisionmakers in their determinations.[333] The most hotly debated topic in juror sentencing deliberations, next to the crime itself, is the issue of the defendant's dangerousness on his return to society.[334] Future dangerousness takes precedence in jury deliberations over any mitigating evidence, such as remorse, mental illness, intelligence, drug/alcohol addiction, and any concern about the defendant's behavior in prison.[335] In fact, dangerousness determinations are part of the explicitly authorized grounds for imposing death in a number of states, and Texas and Oregon even require them.[336] But whether or not statutes permit or require future dangerousness, and whether or not prosecutors even mention it or present it as evidence in the penalty phase of the case, it remains the major focus of the factfinders.[337] This is true regardless of the presence or absence of expert testimony about future dangerousness in the hearing.[338]

---

331. This issue is the one of fit.

332. Albertson, *supra* note 292, at 20.

333. 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." *See* BENCHBOOK, *supra* note 34, at 49 (noting that courts are unwilling to exclude evidence of future dangerousness totally because it is superior to the alternatives).

334. *See* Blume et al., *supra* note 214, at 398–99 (observing, on the basis of interviews with over one hundred capital jurors, that "future dangerousness is in 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").

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

336. *See* OR. REV. STAT. § 163.150(b) (1999) (future dangerousness is statutory aggravating factor); TEX. CRIM. PROC. CODE ANN. art. 37.071 § 2(b) (Vernon 2001) (future dangerousness determinations are authorized ground for death penalty). Idaho, Oklahoma, Virginia, and Wyoming also have future dangerousness as a statutory aggravating factor. IDAHO CODE § 19-2515 (k) (Michie Supp. 2001); OKLA. STAT. ANN. tit. 21, § 701.12 (West 1983); VA. CODE ANN. § 19.2-264.2 (Michie 2000); WYO. STAT. ANN. § 6-2-102(e) (Michie 2001).

337. *See, e.g.*, Garvey, *supra* note 132, at 1560 (citing studies emphasizing "the pervasive role future dangerousness [testimony] plays in and on the minds of capital sentencing jurors").

338. *See* Blume et al., *supra* note 214, 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 defen-

Jurors often believe—incorrectly—that the law requires a death sentence upon a showing of the defendant's future dangerousness.[339] Most people do not believe that defendants sentenced to life actually must spend the rest of their lives in prison.[340] They fear a dangerous person's release into the community; this fear is a highly motivating factor in choosing between death and a life sentence.[341]

A number of factors about jury deliberations increase the likelihood that jurors will overestimate the threat of future violence. The most important of these factors is lack of objective information about such predictions.[342] Base rate errors plague human decisionmaking in general, so there is no reason to suppose that capital jurors are any different. Moreover, jurors seldom receive information on the true rate of violent recidivism among murderers released from prison.[343] Interviews with capital jurors revealed that jurors who sentenced the defendant to death had median estimates of 85% that the defendant would commit a violent crime in the future and a 50% median estimate that the defendant would commit another murder if the defendant had only received a life sentence.[344] Yet empirical studies demonstrate a counter-intuitive

---

dant'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 for jurors to be informed of a death penalty alternative if the prosecution alleges future danger as an aggravating circumstance and the alternative is life in prison without parole. *See* Simmons v. South Carolina, 512 U.S. 154, 169 (1994) (ruling that the court should inform a jury of a defendant's ineligibility for parole when the defendant's future dangerousness is at issue); *see also* Kelly v. South Carolina, 534 U.S. 246, 248 (2002) (reiterating the Court's earlier holding in *Simmons*); *cf.* Garvey, *supra* note 132, at 1560 (observing that "[f]uture dangerousness appears to be one of the primary determinants of capital-sentencing outcomes").

339. *See* Luginbuhl & Howe, *supra* note 286, at 1174 (providing that 43% of surveyed jurors believed that the law required a death sentence upon a showing of future dangerousness).

340. *See Simmons*, 512 U.S. at 159 (citing South Carolina survey showing that only 7.1% of all jury-eligible adults surveyed believed that a life sentence meant that the defendant would be required to spend the rest of his life in prison).

341. *See, e.g., id.* (noting a 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).

342. *See* Sorensen & Pilgrim, *supra* note 122, at 1254 (noting that a lack of objective information about future dangerousness predictions is one of the prime reasons that jurors overestimate the defendant's threat of future violence).

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

344. *Id.* at 1269.

decreased base rate for violence among capital commutees in prison.[345] In one study of 188 death-sentenced prisoners whose sentences the courts commuted after *Furman*, only one killed again, and only six committed violent offenses in the more than five years after their release.[346] In studies of capital commutees paroled into the community, twenty percent returned to prison, but only eight to ten percent committed new felonies.[347] Unless jurors receive information about population base rates and how they should use them in making their decisions, risk estimates amount to little more than speculation.[348]

In addition, jurors consistently underestimate the number of years that a prisoner must serve 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. One well-established principle of criminology is that violence and criminal activity decrease with age.[349] 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.[350]

These factors provide some reason to believe that probabilistic estimates would be helpful to the jury. As long as violence risk assessment is not presented as a yes/no dichotomy, but as a probabilistic assessment, it may aid decisionmaking. Risk estimates are uncertain, and the base rate of serious violence among capital offenders is quite low. In order to be helpful, the expert needs to educate the jury in a scientifically sound manner, including explicitly stating the statistical basis for the opinion.[351]

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 that the VRAG placed in the "high risk" category, for example, only 55% actually committed violent acts upon release. In other words, 45% 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. Because of the centrality of the dangerousness determination to juror deliberations, it would help to the jury to have information

345. Cunningham & Reidy, *supra* note 40, at 23 (citing studies).

346. Sorensen & Pilgrim, *supra* note 122, at 1254-55.

347. Cunningham & Reidy, *supra* note 40, at 24 (citing studies and noting that convicted defendants must now serve longer minimum sentences before parole, making the defendants in the studies much younger than current defendants will be when parole eligible).

348. *Id.* at 24–25.

349. *Id.* at 31.

350. *Id.* (citing studies).

351. *Id.* at 36-38 (advocating that experts limit their testimony to predominantly statistical analyses to avoid going beyond the limits of their scientific expertise).

relating to it. It should, however, be the best information available. Currently, actuarial instruments offer the most accurate way of making such predictions, but the expert must carefully explain the limits of such testimony.[352]

VII. Conclusion

Judges are the relevance gatekeepers for good reason. They have training in critical thinking, they are accountable to superior courts and to legal commentators, and they receive regular feedback regarding how well they have followed their procedures. They should not abdicate this important responsibility when it comes to expert testimony. If the goal of a justice system is accurate determinations, limiting evidence to relevant information is sound cognitive practice.

Rational truth-seeking is the goal of the rule of law also, and thus limiting testimony to what is relevant is a basic notion of procedural fairness. Even expert testimony must abide by this stricture, and therefore courts should limit it to what is relevant, which means that it must have a sound scientific basis. Gatekeeping for scientific validity of expert testimony in capital sentencing proceedings is therefore basic to the rule of law, with its goal of rational truth-seeking.

A decision as important as a death sentence simply cannot be based on bunkum. No civilized country, much less one that prides itself on constitutional principles of due process, should tolerate expert witnesses confusing the jury with wholly unscientific assertions. So, in addition to being sound cognitive practice and mandated by the rule of law, judicial gatekeeping to prevent jury confusion is a minimum for fundamental fairness. Clinical predictions of future dangerousness cannot meet these standards. Actuarial testimony can barely squeak through. Predicting violence, like predicting the weather, is at best subject to a large margin of error. Nonetheless, the structured reasoning process of actuarial risk analysis is far preferable to the ad hoc judgments of clinical predictions. In a system that strives for justice, the least that one can expect is that judges will evaluate expert testimony that may result in a determination of death with as much care as they routinely use to scrutinize expert civil testimony.

Judges who evade the scientific issues presented by expert testimony perpetrate injustice. Providing the jury with misleading and unscientific evidence violates fundamental rule of law principles and due process and is

---

352. *See, e.g.*, Monahan & Steadman, *supra* note 296, at 935-36 (explaining that predictions of the risk of future violence should be modeled explicitly on weather predictions, with all of their qualifiers and uncertainties).

intellectually indefensible. The consequences of misleading the jury in a death sentencing determination are severe, not only for the defendant, but also for a society that values justice and aspires to rationality. The analytical framework provided in this Article aims to enable judges to make better admissibility determinations and to enable juries to reach more reasoned decisions. At stake is the credibility of the judiciary and the proper functioning of the judicial system.