PROVING THE FUTURE IN CRIMINAL CASES

CHRISTOPHER SLOBOGIN*

ABSTRACT

Expert opinion about dangerousness—the risk of reoffending—is commonly introduced at sentencing, criminal commitment proceedings, and some types of pretrial detention hearings. This Essay argues that the rules governing the admissibility of scientific evidence should apply to this testimony and that, on that assumption, such evidence must be (1) "material" (logically relevant, empirically generalizable, and epistemologically germane), (2) "probative" (a measure of accuracy, which is more stringent when the evidence is from an expert), (3) helpful to the factfinder (through promoting "incremental validity"), and (4) presented in a non-prejudicial manner (i.e., in a way that minimizes the possibility it will be misused or misinterpreted). Application of these rules to expert testimony about risk would have significant implications not only for its admissibility in criminal cases but also for the way that testimony is expressed, the law governing dangerousness, and the methods used to assess it.

* Milton Underwood Professor of Law. For their feedback on this Essay, the author would like to thank the participants in the North Sea Group Seminar (based in the Netherlands) and the audience at the Evidence Law panel at the January 2025 American Association of Law Schools meeting in San Francisco. The research for this paper was supported in part by funding from the Weaver Family Program in Law, Brain Sciences, and Behavior.

CONTENTS

INTRODUCTION

Proving the future is always difficult. It is especially so with respect to future offending, where attempts to predict can easily fall prey to miscalculation, bias, and quackery.[1] Yet in numerous settings—in particular, at sentencing, bail hearings, and civil and criminal commitment proceedings—the law insists on asking experts whether particular criminal defendants are "dangerous" or "high risk."[2] This Essay argues that, if courts continue to demand such testimony, evidentiary guardrails must be established. More specifically, this Essay argues that, to be admissible, expert testimony about risk must be (1) "material" (logically relevant, empirically generalizable, and epistemologically germane), (2) "probative" (a measure of accuracy, which is more stringent when the evidence is from an expert), (3) helpful to the factfinder (through promoting "incremental validity"), and (4) presented in a non-prejudicial manner (i.e., in a way that minimizes the possibility it will be misused or misinterpreted).

These rules should be followed even if, as is true in most of the settings in which risk is considered,[3] judges take on the factfinding role. It is often assumed that, in bench proceedings, the rules of evidence need not be vigorously enforced, because judges, unlike jurors, can take flaws in the evidence into account.[4] In part for that reason, the evidence rules that routinely govern at trial are often ignored in pretrial, sentencing, and commitment proceedings, depending on the jurisdiction and the type of crime involved.[5] This stance often extends to the rules governing expert testimony, including the special reliability requirements imposed by the U.S. Supreme Court's holdings in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[6] and its progeny.[7] Further, for better or

[1] *See* Ana M. Otero, *The Death of Fairness: Texas's Future Dangerousness Revisited*, 4 U. DENV. CRIM. L. REV. 1, 30-33 (2014) (detailing incompetent testimony of psychiatrists at Texas death penalty proceedings, including testimony of one expert who was expelled from the American Psychiatric Association as a result).

[2] *See id.* at 2 & nn.3 & 5.

[3] Defendants have a constitutional right to jury determination of sentence in capital cases. *See* Ring v. Arizona, 536 U.S. 584, 607-08 (2002). Additionally, many states provide for juries in sex offender commitments. *See, e.g.*, KAN. STAT. ANN. § 59-29a07 (West 2025). But judges find facts in most pretrial detention, commitment, and sentencing proceedings.

[4] *See, e.g.*, Maggie Wittlin, *Binding Hercules: A Proposal for Bench Trials*, 76 VAND. L. REV. 1735, 1738 (2023).

[5] *See, e.g.*, United States v. Fields, 483 F.3d 313, 338 (5th Cir. 2007) (allowing testimonial hearsay to be admitted at capital sentencing proceedings); 18 U.S.C. § 3142(f)(2)(B) (2018) (highlighting that rules of admissibility of evidence "do not apply" at federal pretrial detention hearings); *cf.* Donald Stone, *There Are Cracks in the Civil Commitment Process: A Practitioner's Recommendations to Patch the System*, 43 FORDHAM URB. L.J. 789, 807 (2016) (noting evidence rules are "loosely applied" at civil commitment hearings).

[6] 509 U.S. 579, 597 (1993).

[7] *See, e.g.*, Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) (holding "that abuse of discretion is the proper standard by which to review a district court's decision to admit or exclude scientific evidence"); Kumho Tire Co. v. Carmichael, 526 U.S. 137, 147 (1999)

worse, the standard of proof in pretrial, commitment, and sentencing proceedings is seldom the proof beyond a reasonable doubt required at a criminal trial.[8] But neither of these realities justifies basing the deprivations of liberty that routinely occur in these proceedings on immaterial, non-probative, unhelpful, or prejudicial information. If the rules of evidence governing the usefulness of expert testimony apply in civil proceedings involving mere damages (*Daubert* was a tort suit), they should apply in criminal and quasi-criminal proceedings as well.

The rules discussed here should, if taken seriously, prevent judges from even considering, much less relying on, certain types of risk evidence that today are routinely admitted. Further, expert testimony that is admitted would be subject to greater scrutiny and would have to be framed differently. Application of traditional evidence rules to expert testimony on risk would also mean that courts, and thus ultimately legislators, would have to think much more carefully about the types of risk that justify deprivations of liberty, and experts would need to devote greater effort to devising research that answers the more demanding questions the law should begin asking. The ultimate impact of imposing an evidentiary framework on expert testimony about risk would be a sea change in the way criminal courts adjudicate risk.

I. APPLICATION OF THE RULES OF EVIDENCE

The Federal Rules of Evidence, widely followed in the states, will be the template for this Essay.[9] Under the Federal Rules, all evidence, lay or expert, must be "relevant" (Rule 402), meaning that it must have a "tendency to make a fact more or less probable than it would be without the evidence" (Rule 401).[10] Additionally, relevant evidence is still inadmissible if its probative value is "substantially outweighed" by the evidence's potential for "unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence" (Rule 403).[11] When the testimony is an opinion from an expert, the probity of the evidence is subject to even greater scrutiny. Following *Daubert*, Rule 702 states that expert opinion must derive from "scientific, technical, or other specialized knowledge," be based on "sufficient facts or data," be "the product of reliable principles and methods," and reflect "a reliable application of the principles and methods to the facts of

(noting that Rule 702 "applies to all expert testimony"); Fields, 483 F.3d at 342 (noting *Daubert* is not applicable at sentencing).

[8] United States v. Redifer, 631 F. App'x 548, 563 (10th Cir. 2015) (refusing to require more than preponderance of evidence standard at sentencing); 18 U.S.C. § 4248(d) (authorizing commitment of sex offenders on "clear and convincing evidence"); 18 U.S.C. § 3142(f) (noting clear and convincing standard of proof applies at pretrial detention hearings).

[9] *See generally* FED. R. EVID.

[10] FED. R. EVID. 401-402.

[11] FED. R. EVID. 403.

the case."[12] Further, the rule provides that the testimony must "help the trier of fact" in deciding the case.[13]

The Federal Rules also set out a few other stipulations unique to expert testimony. Rule 703 provides that experts may rely on otherwise inadmissible information if "experts in the particular field would reasonably rely" on it, but it cautions that this information should not be disclosed to the factfinder unless its probative value would "substantially outweigh[]" its "prejudicial effect."[14] Rule 704 states that expert testimony may "embrace[]" the "ultimate issue" in the case as long as it is not "about whether the defendant did or did not have a mental state or condition that constitutes an element of the crime charged or of a defense."[15] Rule 705 states that experts "may be required" to disclose facts or data underlying their opinion on cross-examination if they are not disclosed on direct,[16] with the caveat from Rule 703 that if the evidence sought to be disclosed is not otherwise admissible, its probative value should substantially outweigh its prejudicial effect. Finally, Rule 706 authorizes the court to appoint its own expert, unless the parties can "show cause" why the appointment should not occur.[17]

If applied vigorously, these rules would have a very significant impact on the admissibility and presentation of expert testimony on risk or dangerousness. The most potent impact would come from Rule 402's requirement that evidence be "relevant," Rule 702's twin requirements that expert testimony be reliable and assist the trier of fact, and Rule 403's admonition that even relevant evidence that is reliable and helpful may be excluded on the ground that it will be misunderstood or misused. In the discussion that follows, the import of these rules is reframed to highlight what I regard to be the four central evidentiary inquiries that arise in connection with testimony about risk.[18]

The first two inquiries are derived from the relevance requirement, which can be divided into two components: a "materiality" inquiry into whether the evidence is logically related to a proposition in the case, and a "probative value" inquiry into whether the evidence is accurate enough to make the proposition "more or less probable." While separating the relevance inquiry into materiality and probative value components is not standard practice, it is crucial in this context, because it recognizes that very accurate information about risk may nonetheless not be logically relevant to any proposition in the case and that, conversely, even evidence about risk that clearly addresses a proposition in a case may nonetheless be irrelevant because it is unreliable. This approach also

[12] FED. R. EVID. 702; *see* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 597 (1993).

[13] FED. R. EVID. 702.

[14] FED. R. EVID. 703.

[15] FED. R. EVID. 704.

[16] FED. R. EVID. 705.

[17] FED. R. EVID. 706.

[18] *See* CHRISTOPHER SLOBOGIN, PROVING THE UNPROVABLE: THE ROLE OF LAW, SCIENCE, AND SPECULATION IN ADJUDICATING CULPABILITY AND DANGEROUSNESS 15 (2006).

allows integration, through the probative value component, of the reliability inquiry required by Rule 702 and the *Daubert* line of cases.

The third central evidentiary requirement recognized in the federal rules is Rule 702's additional stipulation that expert testimony assist the factfinder. This inquiry is separate from the reliability requirement, because even "relevant" (material and probative) expert testimony may be inadmissible if it does not add to what the factfinder could figure out for itself. And finally, even material, probative, and helpful testimony may nonetheless be presented in such a misleading or confusing manner that it should not be admissible.[19]

All four of these admissibility requirements are discussed below. In the course of that discussion, the manner in which Rules 703, 704, 705, and 706 apply to expert testimony about risk should also become clear.

II. MATERIALITY

In the risk assessment setting, the first, and conceptually most difficult, evidentiary issue concerns the materiality of opinion testimony. To see why, a brief description of the various types of expert testimony about risk of reoffending is necessary. Such testimony comes in three basic forms: clinical, structured professional judgment, and actuarial.[20] Clinical evaluation of risk is the most unstructured. It does not limit the historical, psychological, or environmental risk factors to be considered nor does it circumscribe the way in which they are combined to reach a conclusion; essentially, clinical evaluators may arrive at their opinions any way they see fit.[21] In contrast, structured professional judgment ("SPJ") limits the factors the evaluator may consider.[22] For instance, one well-known SPJ approach examines twenty historical, clinical, and management factors, no more and no less; at the same time, it still leaves up to the evaluator how these factors affect any final opinion.[23] Distinguished from both of these risk assessment techniques is actuarial assessment, which both limits the factors to be considered and structures the weights they are assigned so that a quantified estimate of risk is produced. For instance, one actuarial instrument relies on twelve risk factors, each associated with a particular number of points, which are then added together to obtain a total score that allows comparison to the recidivism rates of individuals with the same or similar

[19] *See id.*

[20] *See* 2 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 212-15 (David L. Faigman et al. eds., 2025).

[21] *Id.* at 180 (noting, as example of clinical prediction, that psychiatrists, psychologists, and judges ultimately end up assessing risk based on their personal experiences).

[22] *Id.*

[23] *See* Kevin S. Douglas & Catherine S. Shaffer, *The Science of and Practice with the HCR-20 V3 (Historical-Clinical-Risk Management-20, Version 3)*, *in* HANDBOOK OF VIOLENCE RISK ASSESSMENT 253, 258-59 (Kevin S. Douglas & Randy K. Otto eds., 2d ed. 2021).

scores.[24] Both the SPJ and actuarial assessment approaches use "risk assessment instruments," but only the latter associates numerical probability estimates with an individual.[25]

On the surface, the results from any of these risk assessment techniques may seem material in a proceeding that makes dangerousness a legally relevant criterion. But a more careful analysis would delve into three separate ways risk assessment testimony, even if accurate, might be excluded on logical irrelevance grounds: when it does not address the relevant legal criteria defining risk (legal materiality); when it is based on data that is not generalizable to the case at hand (empirical materiality); and when it is based on risk posed by other people (epistemological materiality). The first two types of materiality are often ignored by the courts; in contrast, the third type needs to be discussed primarily because some courts have *exaggerated* its importance.

A. *Legal Materiality*

Of course, one could argue that expert testimony about risk should never be admissible because dangerousness is not a legitimate legal criterion, especially at sentencing proceedings, where punishment is the goal and thus backward-looking retributive factors might be thought to control. I will not address that argument here, although I have elsewhere.[26] This Essay will assume that risk is a legally legitimate justification for depriving people of liberty at pretrial, commitment, and sentencing proceedings.

Adopting that assumption does not make all testimony about risk legally material, however. Rather, such testimony must address at least three legal issues: the harmful outcome predicted, the period during which that outcome is predicted to occur, and the ways in which the harmful conduct can be prevented.

The requirement that risk testimony focus on the legally relevant outcome is particularly important. Depending on the context, testimony that a given individual will "reoffend," without more, may well be irrelevant. For instance, capital sentencing statutes that make dangerousness an aggravating factor require that the risk posed by the offender involve "acts of violence."[27] Commitment under sexual predator statutes requires proof of a risk for committing "predatory acts of sexual violence."[28] Many civil commitment

[24] *See* N. Zoe Hilton, Marnie E. Rice, Grant T. Harris, Brian Judd & Vernon L. Quinsey, *Actuarial Guides for Appraising the Risk of Violent Reoffending Among General Offenders, Sex Offenders, and Domestic Assaulters*, *in* HANDBOOK OF VIOLENCE RISK ASSESSMENT, *supra* note 23, at 131, 133-34.

[25] *Id.* at 131, 139.

[26] *See* CHRISTOPHER SLOBOGIN, JUST ALGORITHMS: USING SCIENCE TO REDUCE INCARCERATION AND INFORM A JURISPRUDENCE OF RISK 124-31 (2021).

[27] Jurek v. Texas, 428 U.S. 262, 267-68 (1976).

[28] Kansas v. Hendricks, 521 U.S. 346, 350 (1997).

statutes require a risk of "substantial harm."[29] Under these types of statutes, testimony about risk must go well beyond a general conclusion about recidivism.

Unfortunately, it often fails to do so. For instance, many actuarial and SPJ instruments are normed on risk for *any* type of crime (including misdemeanors) or on a definition of violent crime that includes simple assaults,[30] acts which may not (and in my opinion should not) meet the threshold required for significant deprivations of liberty. While *clinical* prediction experts might be able to address this concern simply by asserting that the individual poses the legally relevant risk, courts should demand more than bald declarations on this point, a subject that this Essay will revisit in the discussion of the probative value inquiry.

The second legal criterion that risk testimony must address is the duration of the risk. Many actuarial instruments provide information about risk over follow-up periods of two years or longer.[31] Yet pretrial detention and many sentences are much shorter,[32] making probability estimates from these assessments legally immaterial. Again, given its relative ambiguity in terms of how conclusions are reached, clinical prediction testimony can finesse this issue but should not be allowed to do so.

Finally, both international and domestic law set out mandates (admittedly, not always followed) that require the government to provide evidence about whether the risk posed by the individual requires incapacitation or can instead be addressed through some less restrictive intervention.[33] Actuarial assessments have difficulty answering this question unless they are focused on risk management as well as risk assessment, which they rarely are; many of them

[29] *See, e.g.*, FLA. STAT. ANN. § 394.467(2)(a)(4) (West 2024).

[30] The developers of the *Violence Risk Appraisal Guide* defined violence in the original validation research to include arrest for simple assault and institutional misconduct. *See* Hilton et al., *supra* note 24, at 133; *see also* GRANT T. HARRIS, MARNIE E. RICE, VERNON L. QUINSEY & CATHERINE A. CORMIER, VIOLENT OFFENDERS: APPRAISING AND MANAGING RISK 122-23, 152 (3d ed. 2015); NORTHPOINTE, PRACTITIONER'S GUIDE TO COMPAS CORE 27 (2015), https://archive.epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASPractionerGuide.pdf [https://perma.cc/Q38P-EUYL] (indicating that under company's actuarial instrument, violent offenses include misdemeanor assault offenses); Kevin S. Douglas et al., *Historical-Clinical-Risk Management-20, Version 3 (HCR-20^{V3}): Development and Overview*, 13 INT'L J. FORENSIC MENTAL HEALTH 93, 100 (2014) (defining violent act as one that includes acts that cause "serious psychological harm").

[31] For instance, the Violence Risk Appraisal Guide provides assessments for seven- and ten-year periods. *See* Hilton et al., *supra* note 24, at 132.

[32] *See* Speedy Trial Act of 1974, 18 U.S.C. § 3161 (mandating that time between arrest and trial should normally not exceed 100 days, although also recognizing several circumstances in which that period may be extended).

[33] *See* BERNADETTE MCSHERRY, MANAGING FEAR: THE LAW AND ETHICS OF PREVENTIVE DETENTION AND RISK ASSESSMENT 178-82 (2013) (describing international jurisprudence); Jackson v. Indiana, 406 U.S. 715, 738 (1972) (requiring "reasonable relation" between purpose of confinement and its nature and duration).

rely primarily on "static" risk factors that are not susceptible to change through treatment.[34] Clinical and SPJ testimony, which are more likely to focus on the types of "dynamic" risk factors that are susceptible to treatment, are much better equipped to address the intervention criterion.[35] Again, however, they should be accompanied by some indicia of reliability on this score, in ways outlined in the discussion of the probative value inquiry.

As should be clear from this brief description, legal materiality depends on the law's definition of dangerousness or risk. To date, legislatures and courts have done a very poor job of providing such a definition. As I have argued elsewhere, the principle of legality demands that legal entities develop a more fine-tuned jurisprudence of risk.[36] Just as they have developed a robust legal definition of blameworthiness that delineates the conduct, circumstances, and mental states that justify punishment, legal policymakers should explicitly address the outcome, durational criteria, and intervention criteria that justify preventive confinement. Courts can hasten that development by taking seriously the legal materiality requirement and demanding that litigators and experts explain how proffered risk testimony addresses these criteria.

If legislators have not already done so, courts should also set the *threshold* for intervention criterion—that is, the level of risk (e.g., "more likely than not," "probability," or a particular percentage) that must be met by the relevant standard of proof (e.g., clear and convincing evidence or preponderance of the evidence) before any intervention may take place. It is important to recognize, however, that this criterion does not affect the materiality inquiry. Say, for instance, that the law requires clear and convincing proof that the individual (or a group with his or her characteristics): (1) will more likely than not (2) commit a violent act (3) within the next five years (4) if not imprisoned. Expert testimony must address the last three factors to be material. But it need not assert that the relevant intervention threshold is met. For example, testimony that an individual poses little or no risk, which will usually be proffered by the defense, might simply mean the government cannot prove its case. Or it might be supplemented by other evidence that allows the government to prevail.

B. *Empirical Materiality*

To be material, expert testimony must not only address the relevant legal criteria but also satisfy empirical demands of generalizability, or what the Supreme Court in *Daubert* called "fit."[37] The concern here is whether the sample

[34] For instance, almost all of the Violence Risk Appraisal Guide's twelve risk factors are "static," (e.g., diagnosis, criminal history, family history, age at initial offense, and victim injury). *See* Hilton et al., *supra* note 24, at 133, 136 ("[O]ptimal, long-term, pre-release violence risk assessment can currently be achieved by relying on a comprehensive set of static predictors . . . .").

[35] *See generally* MODERN SCIENTIFIC EVIDENCE, *supra* note 20, at 216-17.

[36] *See* SLOBOGIN, *supra* note 26, at 138-40.

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

on which the expert relies in reaching a conclusion is appropriately focused. Clinical or SPJ experts might find this requirement perplexing, because they view their "sample" as the defendant and the defendant alone; thus, they might respond that of course they have focused on the appropriate sample. This intuition would be correct if all the expert did was offer what I, along with David Faigman and John Monahan, have called "diagnostic testimony."[38] This is testimony that describes how a particular person has acted in the past or is acting in the present; for instance, diagnostic testimony could be a statement that a particular defendant has reacted violently to slurs on his manhood or fondled young boys when alone with them, facts from which one might infer that the defendant is likely to do so again under similar circumstances.[39] This type of testimony is akin to character evidence and may not require an expert opinion at all.

Most expert testimony about risk, however, is what we called "framework" testimony,[40] which is based on general scientific or specialized knowledge—for instance, testimony that people who have certain biological traits, psychological characteristics, or family history (which the defendant has) tend to respond aggressively when their manhood is insulted or to respond sexually when alone with a young boy. When an expert is conveying framework testimony, the "sample" for the opinion is much broader than a particular person and is derived either from research by others or from the evaluator's own research, observations, or experience.[41]

The empirical materiality requirement dictates that the sample "fit" the defendant to the extent feasible.[42] If a psychiatrist has only been trained about and only evaluated people with serious mental illness in emergency commitment settings, or a psychologist relies solely on an SPJ instrument normed on those types of people, their background sample is probably not generalizable to, for instance, a sex offender with a personality disorder. Bad sampling fit is even more likely to be apparent (or, put another way, less easily hidden) when the expert relies on an actuarial instrument, the validation sample for which is (or at least can be) known. Thus, for instance, an instrument validated on a population of seriously mentally ill people would not be empirically material in a sex offender proceeding.

In short, to be empirically material, the data or experiences on which the risk assessment is based should be closely aligned to factors relevant to the

[38] David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417, 443-44 (2014).

[39] *See id.* at 443-46.

[40] *Id.* at 441-43.

[41] *See id.* at 453-54.

[42] *See Daubert*, 509 U.S. at 591 (explaining that proper "fit" requires "expert testimony" to be "sufficiently tied to the facts of the case" to aid the trier of fact (quoting United States v. Downing, 753 F.2d 1224, 1242 (3d Cir. 1985))).

defendant's case. As Min Yang and his colleagues have stated, samples should be similar to the defendant in terms of:

> [D]emographic characteristics (e.g., age, gender, socioeconomic status, ethnicity), level and type of past violence (e.g., criminal histories, sexual vs. nonsexual offenders), psychiatric diagnosis (e.g., presence of personality disorder, psychosis), intervention received (e.g., treated vs. untreated), the specific criterion being predicted (e.g., violent vs. nonviolent behavior or different types of violent behavior), environmental setting (e.g., clients residing in institutions vs. the community), countries of origin of the research, and so forth.[43]

Obviously, some line drawing must occur here; finding a sufficiently large sample with most of the defendant's salient demographic, historical, and environmental features would be impossible. The important point to recognize is that expert testimony based on experience with, or on an actuarial instrument validated on, people who differ significantly from the defendant demographically, geographically, or psychologically should not be considered material, no matter how much experience the expert has or how "accurate" the instrument was shown to be on its validation sample.

In *State v. Loomis*,[44] the Wisconsin Supreme Court recognized the importance of empirical materiality but did not require it.[45] *Loomis* considered the permissibility of using the results of an actuarial risk assessment instrument to inform sentencing. The court alertly warned trial courts that, in using the instrument in future cases, they should take into consideration the fact that it was not yet validated on a Wisconsin population.[46] But not only did the court still permit judges to use the instrument, it also upheld the trade secret claim of the instrument's developer, thereby limiting the ability of litigants and others to evaluate its empirical materiality by hiding the characteristics of the population on which the instrument had been normed[47] (a holding which, it is argued below, should be subject to challenge under Rule 703).[48]

C. *Epistemological Materiality*

Many judges have concluded that, even if it is legally and empirically material and highly accurate at what it purports to describe, framework testimony about risk (to the effect that people like the defendant are, or are not, likely to reoffend) is immaterial, because it is based on information deduced from people other than

---

[43] Min Yang, Stephen C. P. Wong & Jeremy Coid, *The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools*, 136 PSYCH. BULL. 740, 741 (2010).

[44] 881 N.W.2d 749 (Wis. 2016).

[45] *See id.* at 770-71.

[46] *Id.* at 769.

[47] *See id.* at 756, 761.

[48] *See infra* notes 111-14 and accompanying text.

the defendant. For instance, Justice Coyne of the Minnesota Supreme Court deemed "statistics concerning the violent behavior of others irrelevant."[49] Courts in Virginia, Tennessee, and Indiana have excluded testimony about group risk on the like ground that it is not "individualized" or "particularized."[50] Members of the U.S. Supreme Court, in the recent decision of *Diaz v. United States*,[51] have waded into the same territory. In that case, the Court held that framework testimony about whether most drug couriers who cross the border know their car is carrying drugs is not "about" the defendant's mental state; therefore, the testimony was not excludable under Federal Rule 704's prohibition of expert opinion "*about* whether the defendant did or did not have a mental state or condition that constitutes an element of the crime charged."[52] That conclusion led the dissent to ask how such testimony is "relevant" in a trial in which the key mental state issue is a particular defendant's knowledge of whether drugs were present.[53]

The majority in *Diaz* has the better of the argument. As should be clear from the foregoing discussion, all framework testimony is, like the testimony in *Diaz*, based on information about people other than the defendant (albeit people who, if the testimony is empirically material, are similar to the defendant). If experts cannot rely on statistics or inferences drawn from the study or observation of others, they would not be able to testify about the typical symptoms of schizophrenia, the factors that affect eyewitness accuracy, the characteristics that make people suggestible to interrogation techniques, or any other phenomenon expressed in terms of general tendencies, probabilities, or likelihoods. Further exposing the illogic of excluding such evidence, in its absence factfinders will undoubtedly engage in their own stereotyped assumptions about the phenomenon in question, assumptions that may well rest on misconceptions that framework experts can correct. Social psychological research has long demonstrated that a crucial way we come to our best (and worst) understandings of unobservable mental states and propensities is through resort to some level of generalization.[54] As discussed in more detail in connection with the avoidance of prejudice evidentiary requirement,[55] experts should frame their generalized testimony about risk in a way that does not mislead the factfinder. But such testimony should not be excluded on the ground that it is nomothetic (group-based) in nature.

---

[49] *In re* Linehan, 518 N.W.2d 609, 616 (Minn. 1994) (Coyne, J., dissenting).

[50] *See, e.g.*, Porter v. Commonwealth, 661 S.E.2d 415, 440-41 (Va. 2008); United States v. Taylor, 583 F. Supp. 2d 923, 940-42 (E.D. Tenn. 2008).

[51] 602 U.S. 526 (2024).

[52] *Id.* at 534 (emphasis added) (quoting FED. R. EVID. 704(b)).

[53] *Id.* at 543-45 (Gorsuch, J., dissenting).

[54] *See, e.g.*, LEE JUSSIM, SOCIAL PERCEPTION AND SOCIAL REALITY: WHY ACCURACY DOMINATES BIAS AND SELF-FULFILLING PROPHECY 186 (2012) (arguing that stereotypes are useful and practical tools when facing new information).

[55] *See infra* note 102-06 and accompanying text.

Thus, while the testimony in the *Diaz* case about the typical drug courier's knowledge was not literally "about" the defendant's knowledge, unless we discard common epistemological techniques it certainly was relevant to deciding what that knowledge was. Concern about expert generalizations is proper. But it should not result in wholesale exclusion of all nomothetic information. Rather, analysis should be directed at the information's empirical materiality (as discussed above), and its probative value, helpfulness and potential for misuse (discussed below).

III. PROBATIVE VALUE

Ensuring that evidence is material is necessary but, of course, not sufficient. In particular, when evidence is proffered as expert testimony, evidence law has insisted on additional indicia of reliability. In some jurisdictions these indicia simply require a showing that the basis of the opinion is accepted in the relevant field.[56] But in most states the testimony must be, in the words of Rule 702, "the product of reliable principles and methods."[57] Reliability is to be measured, per the *Daubert* decision, by the results of scientific testing or other means of verification, including error rates, peer review, and, last and apparently of least importance, general acceptance in the relevant field.[58]

These additional hurdles to the usual demand that evidence be probative are imposed on expert testimony largely out of concern that lay factfinders will not be able to discern its weaknesses but nonetheless give it substantial weight in reaching a decision.[59] As noted earlier, in the proceedings at issue here, a lay jury is almost never involved. But judges who are legally (but often not otherwise) trained may also have difficulty evaluating scientific evidence.[60] At the same time, the law assumes that they can do so, because they are in charge of deciding whether such evidence is admissible.[61] Given the complexity of evaluating the reliability—or what social scientists would call the validity—of risk assessment techniques, this conundrum is best handled in this context by offloading much of the inquiry to entities other than individual judges.

More specifically, the validity of each type of risk assessment technique would best be decided not by individual judges in individual cases but on a

[56] Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

[57] FED. R. EVID. 702; *see, e.g.*, MASS. R. EVID. 702(c).

[58] Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593-94 (1993).

[59] *Id.* at 595-96.

[60] Valerie P. Hans, *Judges, Juries, and Scientific Evidence*, 16 J.L. & POL'Y 19, 25 (2007) (reporting survey where "judges showed very good understanding of peer review, publication, and general acceptance" but "had more trouble with the concepts of falsifiability and error rate," as well as other research finding that judges are "susceptible, as lay persons are, to various cognitive processing errors and biases, which in turn could compromise their ability to make sound inferences from scientific and statistical evidence").

[61] Gen. Elec. Co. v. Joiner, 522 U.S. 136, 142 (1997) (describing "gatekeeper" role of trial judge).

jurisdiction-wide basis, ideally by an independent body that has access to the relevant research. Admittedly, this proposal departs from the traditional process for vetting expert testimony; the usual assumption, as the Supreme Court itself put it in the context of expert testimony about risk, is that the adversarial system can help the factfinder—judge or jury—"separate the wheat from the chaff."[62] But that is unlikely in this setting for several reasons. First, attorneys in the types of cases at issue here (usually involving indigent defendants) will not have significant resources to expend on expert consultants.[63] Second, an adversarial process that relies on direct and cross-examination of opposing experts is not well-suited to bringing out complicated, nuanced scientific information, especially framework evidence that is based on studies or observations about which even the experts may not be aware.[64] Third, because judges in an adversarial criminal proceeding typically only see two experts, one for each side, they easily succumb to the impression that each side's evidence is plausible even when the testimony of one of the experts is entirely unrepresentative of the scientific consensus. This phenomenon, which can afflict even highly resourced cases, has been referred to as the "99:1 problem," a term meant to convey the fact that in an adversarial system one maverick expert often gets the same exposure as an expert whose views are backed by the rest of the field.[65]

Thus, ideally, rather than requiring the presiding judge to engage, in each case, in time-consuming, often redundant evaluation of the accuracy of a particular risk assessment technique, a legislative, executive branch, or appellate body should do so.[66] Once so vetted, the technique should be considered valid throughout the jurisdiction. Failing adoption of that process, the relevant research should at least be presented by an expert appointed by the court, under Rule 706, rather than by experts chosen by each party (although the parties could still challenge the court's expert through cross-examination and rebuttal evidence). In either case, an entity that is independent of the instrument developer (preferably a university) should aid in assessment of the relevant research.

Such research is most likely to exist in connection with SPJ and actuarial instruments. If instead, the expert is relying on clinical prediction that uses no researchable tool, assessing the validity of the technique is much more difficult,

[62] Barefoot v. Estelle, 463 U.S. 880, 899 n.7 (1983).

[63] JoAnne A. Epps & Kevin Todorow, *Re*fry*ed Forensics: Screening Expert Testimony in Criminal Cases Through* Frye *Plus Reliability*, 48 SETON HALL L. REV. 1161, 1177 (2018) ("Criminal defendants . . . often cannot afford experts. And the criminal justice system generally does not afford them the means to procure experts.").

[64] Andrew W. Jurs, *Balancing Legal Process with Scientific Expertise: Expert Witness Methodology in Five Nations and Suggestions for Reform of Post-*Daubert *U.S. Reliability Determinations*, 95 MARQ. L. REV. 1329, 1343 (2012) ("[C]ross-examination often results in formulaic attack unrelated to the evidence in question.").

[65] Rebecca Haw, *Adversarial Economics in Antitrust Litigation: Losing Academic Consensus in the Battle of the Experts*, 106 NW. U. L. REV. 1261, 1268-70 (2012).

[66] *See* SLOBOGIN, *supra* note 26, at 84-85, 113-15.

because clinicians usually do not (and often cannot) follow up on their predictions. Thus, at most, studies about clinical risk assessment would have to survey the clinical prediction field as a whole, not the validity of any particular clinician's methodology.

With all of this in mind, information about the following indicia of a risk assessment technique's probative value ought to be evaluated: discriminant validity (the ability to distinguish high-risk individuals from low-risk individuals); calibration validity (the ability to associate groups of individuals with risk probabilities); current validity (the ability to give an accurate account of risk under present conditions); general reliability (the ability to make consistent decisions across like cases), and case-specific reliability (the extent to which the expert in the case at hand performed an adequate evaluation).[67] Again, the first four indicia of accuracy would be evaluated on a jurisdiction-wide basis. Only the last aspect of probative value should be evaluated on a case-by-case basis by the judge.

A. *Discrimination*

Given the low base rate of violent crime, if the sole goal is to be correct most of the time, the approach most likely to provide the highest accuracy rate is to predict that no one will recidivate. Depending on the group in question (e.g., people subject to civil commitment, insanity acquittees, sex offenders, or the general prison population), that prediction rule would be wrong only between 5% to 20% of the time with respect to violent offending.[68] But assuming the goal is to identify the highest risk offenders, this blunderbuss mode of decision-making would be useless to the legal system. Instead, the proffered risk assessment technique should have some ability to differentiate high- and low-risk offenders.

Typically, this ability is measured by plotting the true positive rate (the rate at which people designated high risk recidivate) against the false positive rate (the rate at which people designated high risk do not recidivate).[69] Ultimately, this calculation produces an area under the curve ("AUC") value that indicates discriminant validity.[70] An AUC of 1.0 means the risk assessment technique is perfectly accurate, whereas an AUC of 0.5 means it does no better than chance.[71] In the latter case, the instrument would have no probative value because, in the

[67] *See id.* at 68-81 (providing detailed description of validity measures).

[68] *See, e.g.*, PATRICK A. LANGAN, ERICA L. SCHMITT & MATTHEW R. DUROSE, U.S. DOJ, NCJ 198281, RECIDIVISM OF SEX OFFENDERS RELEASED FROM PRISON IN 1994, at 1 (2003) (identifying recidivism rates for serious crime in sex offenders).

[69] *See* Douglas Mossman, *Evaluating Risk Assessments Using Receiver Operating Characteristic Analysis: Rationale, Advantages, Insights, and Limitations*, 31 BEHAV. SCIS. & L. 23, 25, 28-29 (2013).

[70] *Id.* at 30-31.

[71] *Id.* at 31.

words of Rule 401, it does not make a proposition in the case "more or less probable."[72]

The AUCs associated with risk assessment techniques, including clinical techniques, generally range between 0.65 and 0.80.[73] The latter figure indicates that, in studies examining the success rate of the technique, 80% of the time a randomly selected recidivist received a higher score than a randomly selected non-recidivist.[74] While AUC values should be verified by an independent entity, courts would have to decide the AUC threshold a given prediction technique must meet for its results to be admissible.

B. *Calibration*

In addition to a reasonable AUC value, a technique ought to be able to associate a group's characteristics with a reasonably accurate estimate of how many people within that group will recidivate. For instance, one actuarial instrument designates six groups that are associated with ascending levels of recidivism, from 9% to 53%; it also provides confidence intervals in connection with each probability estimate.[75] Lacking this type of information, a court must engage in considerable speculation in making the normative judgment about the risk level that justifies pretrial detention, commitment, or an enhanced sentence.

Calibration data are generally only available with actuarial instruments. SPJ assessments are structured with respect to the risk factors considered, but they leave to the evaluator the decision about whether a person poses a high, medium, or low risk and what those terms mean.[76] Similarly, clinical assessments might conclude that a person poses a high or low risk, but either fail to define those terms or base their conclusions on experience and intuition rather than hard data.[77] The problem in the latter two situations should be apparent from a Rule 702/*Daubert* perspective. Other than general findings suggesting that such predictions are wrong as often as they are right,[78] there are often no "error rates" or other clear indicia of validity that can help judges or other experts assess the

---

[72] FED. R. EVID. 401.

[73] Mossman, *supra* note 69, at 34.

[74] *Id.*

[75] *See Release Conditions Matrix*, ADVANCING PRETRIAL POL'Y & RSCH., https://advancingpretrial.org/psa/factors/release-conditions-matrix [https://perma.cc/N8YX-Z6AG] (last visited May 10, 2025) (correlating Public Safety Assessment scores with likelihood of new pretrial arrest). For confidence interval information, see Matthew DeMichele, Peter Baumgartner, Michael Wenger, Kelle Barrick & Megan Comfort, *Public Safety Assessment: Predictive Utility and Differential Prediction by Race in Kentucky*, 19 CRIMINOLOGY & PUB. POL'Y 409 (2020).

[76] *See* Douglas & Shaffer, *supra* note 23, at 257-58.

[77] *See* MODERN SCIENTIFIC EVIDENCE, *supra* note 20, at 213.

[78] *See, e.g.*, Randy K. Otto, *On the Ability of Mental Health Professionals to "Predict Dangerousness": A Commentary on Interpretations of the "Dangerousness" Literature*, 18 LAW & PSYCH. REV. 43, 63 (1994).

probative value of testimony that someone is "high" or "low" risk. Courts will have to decide whether the relative absence of such information requires them to exclude such testimony or at least accord it less weight.

C. *Currency*

The discriminant and calibration validity of an instrument, even one normed on the jurisdiction in question, can change dramatically over time for a number of reasons. For instance, probability estimates may need adjustment if the make-up of the population in the jurisdiction changes significantly, the jurisdiction's crime, arrest, or conviction rates go up or down, the jurisdiction's policing practices change (with fewer or more people arrested), or the jurisdiction begins implementing innovative alternatives to prison that can reduce risk.[79] All of these factors can affect reoffending rates, so that an instrument validated before the changes take place might estimate a different level of risk than one that was validated after they occur.[80]

Ideally, therefore, periodic audits would be carried out (again, preferably by independent entities) to ensure any risk assessment technique remains sufficiently valid.[81] Periodic auditing is most obviously needed for actuarial instruments, but SPJ and clinical prediction can also be distorted by developments over time. A clinician who evaluates risk based on "years of experience" may be completely out of touch with the newest knowledge about the most potent correlates of violence or the types of people who commit crimes (for example, women are much more likely to be convicted for violent crime today compared to two decades ago).[82] An SPJ or actuarial instrument validated fifteen years ago might pose queries that can lead to invalid conclusions because, for instance, changes in the relevant population have also changed the types of risk factors that are most predictive.

---

[79] *See* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1793-95 (2018) (explaining how changes in law, policy, and population affect risk of failure to appear in court and call into question risk models based on pre-change data).

[80] *See id.*

[81] *See* Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CALIF. L. REV. 439, 489-90 (2020) (explaining need for ongoing, independent auditing of risk models given possibility that they will become outdated due to changes in criminal offense patterns or innovations in pretrial supervision).

[82] *See* LAWRENCE A. GREENFELD & TRACY L. SNELL, U.S. DOJ, NCJ 175688, WOMEN OFFENDERS 5-6 (rev. 2000) (noting increase in convictions of women for violent offenses from 1990 to 2000).

D. *Reliability*

To social scientists, reliability means consistency or repeatability.[83] In the risk assessment setting, a reliability requirement has two separate implications. First, risk assessments between cases should be reliable; experts should reach similar conclusions with respect to people with similar risk factors.[84] This type of reliability is much harder to achieve with individual clinical and SPJ assessments, which can vary appreciably between evaluators, than with actuarial assessments, which are conducted using the same instrument throughout the jurisdiction. However, even with actuarial instruments, inter-rater agreement can vary significantly, especially if the risk factors (e.g., particular diagnoses, definitions of violent crime) are subject to variable interpretations.

A second type of reliability focuses on whether the evaluator, in the words of Rule 702, engaged in "a reliable application of the principles and methods to the facts of the case."[85] This inquiry raises questions such as: Does the defendant have the risk factors (e.g., arrests, convictions, diagnoses) the expert says the defendant has? If an actuarial instrument was used, did the expert score it properly? Did the evaluator make adjustments to the actuarial result and, if so, what were they and why?

This last question is especially important to answer. Because actuarial instruments rely on a limited set of factors, a litigant might object that, in the language of Rule 702, it is not based on "sufficient facts or data";[86] another way of putting this claim is that actuarial assessments are not sufficiently "individualized."[87] Reacting to this concern, an evaluator might adjust the probability estimate based on other factors related to the defendant. While in concept this "adjusted actuarial" approach makes sense, research has shown that such adjustments often detract from accuracy, because they are sometimes based on speculation rather than research, or because they "double-count" risk factors (for instance, criminal history) that are already considered in the algorithm.[88] Thus, this type of assessment should be permitted only if it can be shown that the proposed adjustments are themselves based on valid research and that they are not derived from factors that were considered (and discarded) during the validation process.

---

[83] *See* JOHN MONAHAN & LAURENS WALKER, SOCIAL SCIENCE IN LAW: CASES AND MATERIALS 71-72 (10th ed. 2022).

[84] *See id.* (noting reliability requires consistent measurements).

[85] FED. R. EVID. 702(d).

[86] FED. R. EVID. 702(b).

[87] *See* Vincent M. Southerland, *The Intersection of Race and Algorithmic Tools in the Criminal Legal System*, 80 MD. L. REV. 487, 552, 554 (2021).

[88] *See* Victoria Angelova, Will S. Dobbie & Crystal Yang, *Algorithmic Recommendations and Human Discretion* 20-21 (Nat'l Bureau of Econ. Rsch., Working Paper No. 31747, 2023); Jean-Pierre Guay & Geneviève Parent, *Broken Legs, Clinical Overrides, and Recidivism Risk: An Analysis of Decisions to Adjust Risk Levels with the LS/CMI*, 45 CRIM. JUST. & BEHAV. 82, 83-84 (2018).

As noted earlier, determinations of whether inter-rater reliability is satisfactory (the first reliability issue described above) are best made by an outside body. Only the determination of whether the evaluation in the particular case is sufficiently reliable (issue two) needs to be, and should be, made by the presiding judge, because that determination is specific to the case at hand. This division of responsibility is also more in tune with the relative competencies of the typical judge and the typical lawyer, who will be much more comfortable double-checking the evaluation process in a particular case than assessing non-case-specific issues such as validity and inter-rater reliability.

E. *The Comparative Accuracy of Risk and Culpability Assessments*

Just as the discussion of materiality avoided any conclusion about how much risk justifies imprisonment or other types of interventions, the present discussion has avoided the ultimate issue of how valid and reliable risk assessments must be to meet the requirements of Rule 702 and *Daubert*. Those difficult issues are for the courts and other legal policymakers to decide (although I have made some suggestions on both scores).[89] The only point made here will be a comparative one, focused on the relative accuracy of culpability assessments, which are usually irrelevant in the pretrial and commitment settings but can easily arise in the sentencing context. Those who argue that risk should not be relevant at sentencing sometimes compare the difficulties of assessing risk (which should be obvious from the foregoing) with the alleged relative ease of assessing culpability, given the fact that, in the latter case, the relevant harm can be proven beyond a reasonable doubt, whereas the relevant harm when risk is the issue has not yet occurred (and might not ever occur).[90]

This would be an apt point if the culpability assessment at issue is whether an individual should be convicted at trial. But the assertion that culpability assessments are more accurate than risk assessments is made in aid of the argument that culpability should be the predominant focus at *sentencing*. Thus, the more appropriate comparator to the accuracy of risk assessments is the accuracy of culpability determinations that are typically made in fashioning criminal dispositions, which often turn not on the harm the defendant caused but on gradations of mental state at the time of the offense and, in more progressive courts, the biological or environmental constraints on the defendant's ability to choose otherwise.[91] To the extent that is so, it is not at all clear that the culpability assessments made at sentencing are any more "accurate" than risk assessments at sentencing. Indeed, in contrast to risk assessment accuracy,

[89] *See* SLOBOGIN, *supra* note 26, at 57-63.

[90] *See* MICHAEL TONRY, DOING JUSTICE, PREVENTING CRIME 21, 160 (2020) (arguing in favor of sentencing based on proportional blameworthiness, in part because "predictions are more often inaccurate than accurate").

[91] *See, e.g.*, ANDREW VON HIRSCH & ANDREW ASHWORTH, PROPORTIONATE SENTENCING: EXPLORING THE PRINCIPLES 143-46 (2005) (explaining that gauging culpability at sentencing involves nuanced evaluation of mental state and excuse at a granular level).

which can be scientifically investigated, the existence, content, and impact on behavior of subjective mental states and external factors as they relate to "desert" are not susceptible to measurement.[92]

IV. HELPFULNESS

Even if expert testimony on risk is material and probative, it may be excluded on the ground that it does not "assist the factfinder" as required by Rule 702, because it consists of a commonsense assessment that laypeople are capable of making on their own. If so, the expert testimony becomes superfluous. The most obvious counter to this contention is that the delineation of risk factors and their relative impact on specific defendants is not something laypeople can easily discern. Additionally, valid risk assessments can overcome the natural, but erroneous, perception that people who have committed violent antisocial acts in the recent past are likely to do so again. In fact, as noted earlier,[93] most such individuals do not reoffend violently, a point that expert testimony can helpfully bring home in a quantified or quasi-quantified way. In short, expert testimony about risk can be helpful—or, in scientific terms, provide information with incremental validity—to the decision-maker.

In a 2018 study, Dressel and Farid purported to expose the fallacy of this reasoning by comparing lay decisions about risk to those reached by a well-known actuarial instrument, the Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS").[94] They found that, in over 1,000 assessments, humans were correct in about 62% of the cases, and the COMPAS was correct in 65% of the cases.[95] A number of commentators have pointed to this study in arguing that actuarial instruments are no more accurate than lay factfinders and thus are of no help in the sense required by Rule 702.[96]

The Dressel and Farid study suffered from one significant problem, however. The human subjects were each shown fifty short vignettes that listed only a few features of the defendant, all of which have a robust statistical relationship with reoffending, and were also immediately told, after each decision, whether they were right or wrong.[97] In effect, this methodology turned the humans into algorithms, because they were limited in the factors they could consider and were "trained" to use certain risk factors and not others.[98] A follow-up study by

[92] *See* SLOBOGIN, *supra* note 18, at 42-48.

[93] *See supra* note 68 and accompanying text.

[94] Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES 1, https://www.science.org/doi/epdf/10.1126/sciadv.aao5580 (last updated Mar. 30, 2018).

[95] *See id.* at 2.

[96] *See, e.g.*, Utsav Bahl, Chad Topaz, Lea Obermüller, Sophie Goldstein & Mira Sneirson, *Algorithms in Judges' Hands: Incarceration and Inequity in Broward County, Florida*, 71 UCLA L. REV. DISCOURSE 246, 248 & n.4 (2024).

[97] *See* SLOBOGIN, *supra* note 26, at 66.

[98] *Id.*

Lin et al. found that when lay participants are not provided this type of guidance and feedback, they do much more poorly than an actuarial instrument, even when they are given the base rate of offending among the population in question.[99] The latter study also found that when the information given the humans was "noisier" (that is, richer in detail than the narrow list of salient factors Dressel and Farid gave their subjects), the humans did barely better than chance, whereas the statistical model had a much higher AUC value.[100] The Lin et al. study suggests that judges, who rarely get feedback about their decisions and are virtually always confronted by considerable noise in pretrial, commitment, and sentencing proceedings, are likely to do much worse on their own than when aided by valid expert testimony about risk. That finding is consistent with voluminous research suggesting that mathematical models are more accurate than unstructured intuitive judgments.[101]

V. PREJUDICE

Material, probative, and helpful expert testimony about risk assessment might still be excluded if, to use Rule 403's language, the "danger" that it will be misused is great.[102] This danger might vary depending on the type of risk testimony. Actuarial-based testimony can be confusing or hard to understand. Clinical and SPJ testimony is less likely to be confusing but, in part for that reason, is more likely to be interpreted by the factfinder as a definitive pronouncement that the defendant will (or will not) offend. This likelihood is a danger because, as even those who rely on clinical and SPJ reasoning acknowledge, "predictions of future offending cannot be achieved, with any degree of confidence, in the individual case."[103] In other words, experts do not have the scientific or specialized knowledge to predict the future of a particular individual with certainty, so any testimony that suggests otherwise could violate Rule 403.

However, the rules of evidence can be deployed to limit these types of dangers. First, experts should not be permitted to assert that a particular defendant will or will not recidivate, nor be able to use language (e.g., "likely to

[99] Zhiyuan "Jerry" Lin, Jongbin Jung, Sharad Goel & Jennifer Skeem, *The Limits of Human Predictions of Recidivism*, SCI. ADVANCES 2-4 (Feb. 14, 2020), https://www.science.org/doi/epdf/10.1126/sciadv.aaz0652.

[100] *Id.* at 5.

[101] Sarah L. Desmarais, Kiersten L. Johnson & Jay P. Singh, *Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings*, 13 PSYCH. SERVS. 206, 206 (2016) ("There is overwhelming evidence that risk assessments completed using structured approaches produce estimates that are more reliable and more accurate than unstructured risk assessments.").

[102] FED. R. EVID. 403.

[103] David J. Cooke & Christine Michie, *Limitations of Diagnostic Precision and Predictive Utility in the Individual Case: A Challenge for Forensic Practice*, 34 LAW & HUM. BEHAV. 259, 272 (2010).

recidivate") that cannot be backed up with empirical information. Although Rule 704 permits testimony that embraces the ultimate issue,[104] Rule 702 requires that all expert testimony stem from "scientific, technical, or other specialized knowledge";[105] a statement that a given individual will or will not recidivate cannot meet this test. Thus, if the expert testimony about risk is based on an actuarial instrument, the opinion should be framed along the following lines: "the defendant shares characteristics A, B, and C with a group, X to Y% of which commit an O type of offense within T period of time if they are not subject to I types of interventions." If instead the testimony is based on an SPJ instrument or is clinical in nature, the testimony should avoid any mention of probabilities (because any such mention would be pure guesswork). In particular, contrary to typical practice among some SPJ experts, the evaluator should be prohibited from saying that a person's risk is "high," "medium," or "low," unless these words are empirically tied to probability estimates or at least compared to some identifiable standard.[106] Instead, these evaluators should be limited to describing the conditions which, given research or past experience, are most likely to enhance or decrease the chance of offending. Only in these ways can experts avoid misrepresenting the nature of their expertise.

The dissent in *Diaz*, the Supreme Court case noted earlier, claimed that these types of limitations on opinion testimony are disingenuous. Testimony that most drug couriers know they are carrying drugs, the dissent argued, is identical to testimony that *Diaz* knew she was carrying drugs.[107] But, in fact, the two statements are quite different. Assuming Diaz told no one about her mental state, only she could know whether she was aware drugs were in her car when she was stopped. But the rest of us can know, at least in theory, what most drug couriers know. Further, while it is not clear how the expert in *Diaz* was able to describe what most drug couriers know (a probative value issue), if research determines that most couriers have this knowledge, communicating that information can be helpful to the factfinder, which otherwise might assume most drugs found in cars at the border were planted there unbeknownst to the courier. Risk assessment testimony, especially when based on actuarial analysis, is even more helpful, because it replaces words like "most," used in the *Diaz* case, with a probability range that is more concrete and probative.

A second means of insuring against the dangers that expert opinion evidence will be misused is to take seriously Rule 705's provision that the facts

[104] FED. R. EVID. 704 ("An opinion is not objectionable just because it embraces an ultimate issue.").

[105] FED. R. EVID. 702.

[106] Nicholas Scurich, *The Case Against Categorical Risk Estimates*, 36 BEHAV. SCIS. & L. 554, 556-57 (2018) (finding that there is "no consensus among risk communicators" on what these terms mean).

[107] *See* Diaz v. United States, 602 U.S. 526, 548 (2024) (Gorsuch, J., dissenting) (arguing that there is "no difference" between "definitive" and "probabilistic" expert opinions).

underlying an opinion can always be exposed through cross-examination.[108] In fact, any good expert should describe these facts during *direct* examination, in the course of explaining how the opinion was reached. There should be only two scenarios where that might not happen.

The first is when, under Rule 703, the judge determines that the probative value of the facts underlying the expert opinion does not significantly outweigh their prejudicial effect.[109] Notably, this formulation reverses the usual balancing analysis under Rule 403, which admits evidence unless its probative value is substantially outweighed by its prejudicial impact. In jurisdictions that require sentencing entities to consider both retributive and risk factors, Rule 703 might, in theory, be triggered when the factors underlying a conclusion that the defendant poses a low risk (e.g., maturity, lack of impulsivity) also indicate *increased* culpability; risk factors are often orthogonal to factors relevant to desert and thus if introduced into evidence could be used by the factfinder for a purpose other than intended.[110] But this type of conflict should rarely, if ever, lead to exclusion of relevant facts. If both risk and culpability are legitimate sentencing criteria, hiding material and probative facts from the factfinder will ensure flawed reasoning on risk, culpability, or both. In any event, this concern is theoretical in those settings where a judge is the factfinder, since the judge will already know the relevant facts or find out about them if an objection under Rule 703 is made.

Litigants are more likely to make a Rule 703 objection when they believe the expert's description of underlying facts would disclose a trade secret, the type of claim mentioned earlier in connection with the *Loomis* case.[111] The contention here is that, when a private company develops a risk assessment instrument, it has a protected intellectual property interest that would be breached if the instrument's risk factors, their weights, and how they are combined are divulged in open court and exposed to its competitors.[112] *Loomis* upheld that argument.[113] But the foregoing discussion about why knowledge of a tool's contents and its validation process is important—both for vetting entities assessing empirical materiality and validity and for judges and lawyers double-checking the adequacy of an expert's evaluation process—counsels the opposite conclusion. Additionally, if the trade secret claim seeks to prevent a prosecution witness from explaining the basis of an opinion, the constitutional right of the accused

[108] FED. R. EVID. 705.

[109] FED. R. EVID. 703.

[110] *See* Megan T. Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 WASH. U. L. REV. 681, 684-86 (2018).

[111] *See supra* notes 44-48 and accompanying text.

[112] *See* State v. Loomis, 881 N.W.2d 749, 761 (Wis. 2016).

[113] *Id.*

to confront witnesses may be implicated and should, at the least, be accommodated through *in camera* review, protective orders, and the like.[114]

CONCLUSION

Expert testimony about risk should be legally and empirically material, meet the validity tests of Rule 702 and *Daubert*, add to the factfinder's knowledge about the defendant's risk, and be presented in a manner that minimizes misuse. These inquiries are more easily pursued when the expert evidence is based on actuarial rather than clinical or SPJ evaluations, given the quantified nature of actuarial assessments. But they should apply regardless of the type of testimony proffered.

[114] *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1409-13 (2018).