

PREVENTIVE JUSTICE: HOW ALGORITHMS, PAROLE BOARDS, AND LIMITING RETRIBUTIVISM COULD END MASS INCARCERATION

*Christopher Slobogin**

A number of states use statistically derived algorithms to provide estimates of the risk of reoffending. In theory, these risk assessment instruments could bring significant benefits. Fewer people—of all ethnicities—would be put in jail prior to trial and in prison after conviction, the duration of sentences would be reduced for low-risk offenders, and treatment resources would be more efficiently allocated. As a result, the capital outlays for prisons and jails would be substantially reduced. The public would continue to be protected from the most dangerous individuals, while lower-risk individuals would be less subject to the criminogenic effects of incarceration and better positioned to build and maintain a life outside of jail or prison that does not involve criminal activity.

Risk assessment instruments cannot fully realize these benefits, however, unless the currently popular determinate sentencing structure that exists in most states is dramatically altered. Today, determinate sentencing states give almost all sentencing power to prosecutors, who in essence fix the sentence range through charging practices, and judges, who decide where within the range the sentence will fall and occasionally select a sentence outside that range. The result is that even an offender who poses a low risk of reoffending will often receive a lengthy sentence of imprisonment.

* For their feedback on this Article, the author would like to thank participants in workshops at Duke, Chicago, Ohio State, Utah, Vanderbilt, Washington Law Schools, and at Oxford University, as well as Melissa Hamilton, John Monahan, Michael O’Hear, and Jennifer Skeem. A small part of this Article appeared as *Limiting Retributivism and Individual Prevention*, in *THE ROUTLEDGE HANDBOOK ON THE PHILOSOPHY AND SCIENCE OF PUNISHMENT* (Farah Focquaert, ed., 2020). A much fuller version of this Article and additional material will appear in a forthcoming book to be published by Cambridge University Press, entitled *JUST ALGORITHMS: USING SCIENCE TO REDUCE INCARCERATION AND INFORM A JURISPRUDENCE OF RISK* (forthcoming 2021).

This Article describes what it calls a preventive justice sentencing regime, which adopts sentence ranges consistent with the offender's desert and then relies on expert parole boards to determine the nature and duration of sentence within this range, based on consideration of individual prevention goals (i.e., incapacitation, specific deterrence, and rehabilitation) as measured through risk assessment instruments. In the course of doing so, it defends risk assessment instruments, which have been subject to a wide range of attacks on accuracy and fairness grounds. A well-constructed system of preventive justice can alleviate many of the inherent tensions between desert and prevention, between deontology and political reality, and between the desire for community input and the allure of expertise. If done properly, it should also significantly reduce prison populations.

TABLE OF CONTENTS

I.	INTRODUCTION.....	99
II.	PREVENTIVE JUSTICE AND LIMITING RETRIBUTIVISM.....	109
	A. <i>The Determination of Sentence Ranges</i>	112
	B. <i>Governing Principles of Preventive Justice</i>	116
	C. <i>Parole as a Constitutional Right</i>	120
III.	THE USEFULNESS OF RISK ASSESSMENT INSTRUMENTS.....	125
	A. <i>The Nature of Risk Assessment Instruments</i>	127
	B. <i>Ensuring the Materiality, or Fit, of RAIs</i>	131
	C. <i>Ensuring the Validity of RAIs</i>	134
	D. <i>The Need for a Presumption in Favor of RAI Results</i>	138
IV.	A DEFENSE OF RISK ASSESSMENT INSTRUMENTS AGAINST INJUSTICE CLAIMS.....	140
	A. <i>The Claim of Egalitarian Injustice</i>	141
	1. <i>Race</i>	145
	2. <i>Sex</i>	150
	3. <i>Other Traits</i>	152
	B. <i>The Claim of Retributive Injustice</i>	154
	C. <i>The Claim of Procedural Injustice</i>	159
	1. <i>Voice</i>	160
	2. <i>Transparency</i>	164
V.	CONCLUSION—THE NEED FOR EXPERIMENTATION.....	168

I. INTRODUCTION

When it comes to criminal justice, change is in the air. From calls to Defund the Police¹ and Abolish Prisons² to the Eight Can't Wait³ and Smart Sentencing⁴ initiatives, a cacophony of proposals for reorienting policing, pretrial detention decision-making, and sentencing are jumping out of academic journals and into mainstream political culture. This Article adds to the reformist hubbub by proposing dramatic changes to our system of punishment. It calls for a system of “preventive justice” that harks back to the days of sentences determined by parole boards, but with two important twists: sentence ranges would be consistent with retributive principles, and release would be required at the expiration of the low end of the range unless the offender is found to pose a high risk for committing violent crime, based on the results of a statistically derived risk assessment tool.

The prescriptions advanced here will undoubtedly strike many would-be reformers as the opposite of a reform agenda—a throwback to worn-out ideas and a dangerous endorsement of flawed, biased, and mechanistic technologies. The burden of this Article is to persuade otherwise.

Meeting that burden begins by emphasizing that a primary goal of preventive justice is to put a significant dent in our incarceration rates and massive prison populations. It is well-known that the imprisonment rate in the United States has skyrocketed since the late 1960s, from the neighborhood of one hundred people per one hundred thousand, to somewhere between five to seven hundred people per one hundred thousand, so that now prisons and jails house well over two million individuals.⁵ Although prison growth has moderated somewhat in the past several years,⁶ the pace of contraction has been

1. See Black Lives Matter Can., *It's Time for a Change*, DEFUND THE POLICE, <https://defundthepolice.org/> (last visited Mar. 15, 2021).

2. See *Landing Page*, ABOLISH PRISONS, <https://abolishprisons.org/> (last visited Mar. 15, 2021).

3. See Campaign Zero, *Landing Page*, #8CANTWAIT, <https://8cantwait.org/> (last visited Mar. 15, 2021).

4. See *Smart Sentencing Coalition*, EMPOWER MO., <https://empowermissouri.org/smart-sentencing-coalition/> (last visited Mar. 15, 2021).

5. Compare ROY WALMSLEY, INST. FOR CRIM. POL'Y RSCH., WORLD PRISON POPULATION LIST 2, 6 tbl.2 (12th ed. 2018), https://www.prisonstudies.org/sites/default/files/resources/downloads/wpppl_12.pdf (listing the U.S. rate as 655 imprisoned per 100,000), with E. ANN CARSON, U.S. DEP'T OF JUST., NCJ 253516, PRISONERS IN 2018, at 9 tbl.5, 10 tbl.6, 2 (2020), <https://www.bjs.gov/content/pub/pdf/p18.pdf> (listing the U.S. rate as 431 imprisoned per 100,000 for residents of all ages and 555 imprisoned per 100,000 for residents age eighteen or older despite also including individuals in “boot camps, halfway houses, treatment facilities, hospitals, local jails, or another state's facilities”).

6. JACOB KANG-BROWN ET AL., VERA INST. OF JUST., PEOPLE IN PRISON IN 2018, at 1–2 (2019), <https://www.vera.org/downloads/publications/people-in->

slow. By one estimate, at the current rate of downturn, it will take until 2101 for the prison population to return to its 1980 level.⁷

A growing number of policymakers and commentators think that postconviction incarceration rates are an acute problem. Those on the left are most concerned about the human cost, not only to suspects and offenders but to their families and their communities. In particular, these critics point to the huge proportion of people of color who are in prison—at a rate roughly six times that of whites⁸—and to the disruption that imprisonment causes to the families and neighborhoods of those who are confined.⁹ More generally, critics

prison-in-2018-updated.pdf (noting a 1.8 percent drop in prisoners, but also noting the drop is largely attributable to reductions in federal prisons and in seven states, at least four of which have “large prison populations,” while “[t]he number of people in prison increased in 19 states”).

7. Nazgol Ghandnoosh, *Minimizing the Maximum: The Case for Shortening All Prison Sentences*, in SMART DECARCERATION: ACHIEVING CRIMINAL JUSTICE TRANSFORMATION IN THE 21ST CENTURY 137, 139 (Matthew W. Epperson & Carrie Pettus-Davis eds., 2017); see also Malcolm C. Young, *Why ‘Tweaking Around the Edges’ Won’t Reduce Mass Incarceration*, CRIME REP. (May 2, 2019), <https://thecrimereport.org/2019/05/02/why-tweaking-around-the-edges-wont-reduce-mass-incarceration/> (predicting that, at best, the American prison population—which does not include jail tallies—will not fall below one million until 2042).

8. MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS 148 (2020) (citing EILEEN POE-YAMAGATA & MICHAEL A. JONES, NAT’L COUNCIL ON CRIME & DELINQ., AND JUSTICE FOR SOME: DIFFERENTIAL TREATMENT OF YOUTH OF COLOR IN THE JUSTICE SYSTEM (2000)). The disparity holds true even when other factors are held constant. See COMM. ON LAW & JUST., NAT’L RSCH. COUNCIL OF THE NAT’L ACADS., THE GROWTH OF INCARCERATION IN THE UNITED STATES: EXPLORING CAUSES AND CONSEQUENCES 5 (Jeremy Travis et al. eds., 2014) (“Among white male high school dropouts born in the late 1970s, about one-third are estimated to have served time in prison by their mid-30s. Yet incarceration rates have reached even higher levels among young black men with little schooling: among black male high school dropouts, about two-thirds have a prison record by that same age—more than twice the rate for their white counterparts.”).

9. See COMM. ON LAW & JUST., NAT’L RSCH. COUNCIL OF THE NAT’L ACADS., *supra* note 8, at 6 (“The partners and children of prisoners are particularly likely to experience adverse outcomes if the men were positively involved with their families prior to incarceration.”); FREDERIC G. REAMER, ON THE PAROLE BOARD: REFLECTIONS ON CRIME, PUNISHMENT, REDEMPTION, AND JUSTICE 257 (2017) (“Incarceration is also strongly correlated with negative social and economic outcomes for former prisoners and their families. . . . Men with a criminal record often experience reduced earnings and employment after prison Fathers’ incarceration and family hardship, including housing insecurity and behavioral problems in children, are strongly related.”); Nicole Lewis & Beatrix Lockwood, *How Families Cope with the Hidden Costs of Incarceration for the Holidays*, N.Y. TIMES (Dec. 20, 2019), <https://www.nytimes.com/2019/12/17/us/incarceration-holidays-family-costs.html> (stating that families spend \$2.9 billion a year on

emphasize the research that suggests that imprisonment is itself criminogenic.¹⁰ Those on the right may have these reactions as well,¹¹ but are probably at least as concerned about the cost of prisons, jails, and correctional staff.¹² The money spent on the carceral state has quadrupled since the 1990s, to upwards of \$80 billion a year,¹³ and the indirect costs of this prison boom have been estimated at over \$500 billion (about 6 percent of the U.S. gross domestic product).¹⁴

commissary accounts and phone calls, have to buy basic hygiene items and other necessities for prisoners, and pay about \$13,000 per prisoner in fines and fees).

10. DAVID ROODMAN, *THE IMPACTS OF INCARCERATION ON CRIME* 77–119 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635864 (analyzing different studies regarding the aftereffects of incarceration); Daniel S. Nagin et al., *Imprisonment and Reoffending*, in 38 *CRIME AND JUSTICE: A REVIEW OF RESEARCH* 115 (Michael Tonry ed., 2009) (evaluating research and concluding that studies mildly point toward a criminogenic effect, though the research is inconclusive); William D. Bales & Alex R. Piquero, *Assessing the Impact of Imprisonment on Recidivism*, 8 *J. EXPERIMENTAL CRIMINOLOGY* 71, 98 (2012) (finding, based on a study in Florida, that offenders released from prison committed more crime than matched offenders released from community programs); Martin H. Pritikin, *Is Prison Increasing Crime?*, 2008 *WIS. L. REV.* 1049, 1082 (“[W]e can tentatively estimate that incarceration causes about 7 percent of total crime: 1 percent because of in-prison crime, 2 percent because of prison-induced recidivism, and 4 percent because of the impact of incarceration on the delinquency of inmates’ children.”). *See generally* Michael Tonry, *Less Imprisonment is No Doubt a Good Thing: More Policing is Not*, 10 *CRIMINOLOGY & PUB. POL’Y.* 137, 137–38 (2011) (“The effects of imprisonment on individual deterrence are most likely perverse; people sent to prison tend to come out worse and more likely to reoffend than if they had received a lesser punishment. . . . [T]entative but not yet conclusive evidence indicates that imprisonment is criminogenic and increases released inmates’ rates of reoffending.”).

11. *See, e.g.*, Press Release, Law Enft Leaders to Reduce Crime & Incarceration, Law Enforcement Leaders Urge President to Back Sentencing Reform, Spurn More Incarceration as Part of Anti-Crime Agenda 1–2 (Feb. 13, 2017), <http://lawenforcementleaders.org/wp-content/uploads/2017/02/LEL-Report-Release.pdf>.

12. *See* DAVID DAGAN & STEVEN E. TELES, *PRISON BREAK: WHY CONSERVATIVES TURNED AGAINST MASS INCARCERATION* xi–xii, 2 (2016); Charlie Savage, *Trend to Lighten Harsh Sentences Catches on in Conservative States*, *N.Y. TIMES* (Aug. 12, 2011), <http://www.nytimes.com/2011/08/13/us/13penal.html>.

13. CHRISTINA HENRICHSON & RUTH DELANEY, *VERA INST. OF JUST., THE PRICE OF PRISONS: WHAT INCARCERATION COSTS TAXPAYERS* 2 (Jules Verdone ed., 2012), https://www.vera.org/downloads/Publications/price-of-prisons-what-incarceration-costs-taxpayers/legacy_downloads/price-of-prisons-updated-version-021914.pdf; Lewis & Lockwood, *supra* note 9 (“The Bureau of Justice Statistics estimates that the United States spends more than \$80 billion each year to keep roughly 2.3 million people behind bars.”).

14. Michael McLaughlin et al., *The Economic Burden of Incarceration in the U.S.* 2, 4, 32 (Concordance Inst. for Advancing Soc. Just., Working Paper No. CI072016, 2016), <https://joinnia.com/wp-content/uploads/2017/02/The-Economic-Burden-of-Incarceration-in-the-US-2016.pdf>.

From a more philosophical perspective, both those who believe punishment should focus on just desert and those who are more utilitarian in orientation favor some degree of decarceration.¹⁵ Whatever the reason, a consensus has built among policymakers that reducing the number of people in prison is a critical objective, at least if it can be done without increasing the danger to the public.¹⁶

The public seems to agree. In a 2006 poll, only 38 percent of respondents said that reducing the prison population was “very important” as a stand-alone goal, while 81 percent stated it was “very important” to ensure that “the punishment fits the crime.”¹⁷ But by 2016, surveys showed that over 80 percent favored reduction of prison populations as a primary goal of the criminal justice system.¹⁸ And when given more context, the public’s attitudes toward incarceration appear to be even more attuned to its negative aspects. In one 2012 PEW poll, 78 percent of respondents stated that it would be acceptable to reduce prison time for low-risk, nonviolent offenders to close budget deficits,¹⁹ and over 80 percent believed that more money should be spent on alternatives to prison for such offenders.²⁰ Well over a majority of respondents in the same poll endorsed the following statement: “It does not matter whether a nonviolent offender is in prison for 18 or 24 [or] 30 months What really matters is that the system does a better job of making sure that when an offender

15. See generally Jessica Kelley & Arthur Rizer, *Keep Calm and Carry on with State Criminal Justice Reform*, 32 FED. SENT’G. REP. 86 (2019) (describing the purposes of criminal justice and the changes implemented by the First Step Act in criminal justice reform); Tonry, *supra* note 10, at 138, 144.

16. See DAGAN & TELES, *supra* note 12, at xi–xiii (discussing the conservative shift toward reducing the prison population); see also Justin McCarthy, *Americans’ Views Shift on Toughness of Justice System*, GALLUP (Oct. 20, 2016), <https://news.gallup.com/poll/196568/americans-views-shift-toughness-justice-system.aspx>; Paul Samuels & Gabrielle De La Guéronnière, Opinion, *Candidates Take Note: Strong Bipartisan Consensus on Criminal Justice Reform*, THE HILL (Dec. 23, 2019), <https://thehill.com/opinion/criminal-justice/475732-candidates-take-note-strong-bipartisan-consensus-on-criminal-justice>.

17. PRINCETON SUR. RSCH. ASSOCS. INT’L FOR THE NAT’L CTR. FOR STATE CTS., THE NCSC SENTENCING ATTITUDES SURVEY 38 (2006).

18. See Holly Harris & Andrew Howard, *Ryan’s Victory Trumps Justice Reform Opponents*, THE HILL (Aug. 15, 2016), <http://thehill.com/blogs/pundits-blog/crime/291500-ryans-victory-trumps-justice-reform-opponents> (“[P]olling data from dozens of states across the country shows overwhelming support across the political and ideological spectrum for criminal justice reform.”); *Voters Want Big Changes in Federal Sentencing, Prison System: Majority Supports Broad Reforms for Drug Offenses, National Poll Finds*, PEW CHARITABLE TRS. (Feb. 12, 2016), <https://www.pewtrusts.org/en/research-and-analysis/articles/2016/02/12/voters-want-changes-in-federal-sentencing-prison-system>.

19. PEW CTR. ON THE STATES, PUBLIC OPINION ON SENTENCING AND CORRECTIONS POLICY IN AMERICA 4 (2012).

20. *Id.* at 1, 4, 7.

does get out, he is less likely to commit another crime.”²¹ In two polls of Wisconsin citizens (a “purple” state) in 2012–2013, a majority of respondents supported early release for both nonviolent *and* violent offenders—at a point halfway through the sentence no less—if the offender “can demonstrate that he is no longer a threat to society.”²²

Several solutions to the mass incarceration problem have been proposed and, in some states, have been implemented. Prominent initiatives include shortening sentences either at the front end or through early release, eliminating mandatory sentencing, creating more alternatives to jail and prison, and supporting funding initiatives prohibiting the use of state prisons for certain categories of offenders, which in California created an incentive for localities to develop cheaper options.²³ More radical proposals—such as those subsumed under the Abolish Prison rubric—are unlikely to go anywhere, but in today’s protest-energized world are more than just an academic pipedream and can at least provide baselines against which to measure more modest reforms.

As both the poll data and the experience of reform states indicate, however, substantial change in incarceration practices will not occur unless politicians, government officials, and the public have reasonable assurances that most of the individuals who are meant to benefit from these types of reforms will not commit new serious felonies. Even advocates for abolishing prisons always express the important caveat that confinement needs to be retained for the “dangerous few.”²⁴ Those focused on civil liberties have an additional concern. They worry that, to the extent these initiatives depend upon the subjective judgments of judges and parole boards, the move toward discretion-based sentences and intermediate dispositions will not be evenly distributed, but rather will disfavor people of color or others who fit certain stereotypes.²⁵

The key thesis of this Article is that risk assessment instruments (“RAIs”)—statistically derived algorithms that estimate the risk of reoffending posed by groups of offenders—can play a significant role

21. *Id.* at 5.

22. Michael O’Hear & Darren Wheelock, *Imprisonment Inertia and Public Attitudes Toward “Truth in Sentencing,”* 2015 BYU L. REV. 257, 276 (2015).

23. For a description of realignment and its impact, see Allen Hopper et al., *Shifting the Paradigm or Shifting the Problem?: The Politics of California’s Criminal Justice Realignment*, 54 SANTA CLARA L. REV. 527, 554–93 (2014).

24. Allegra M. McLeod, *Prison Abolition and Grounded Justice*, 62 UCLA L. REV. 1156, 1168 (2015) (“Who and how many are the dangerous few? The answer to this question is by no means self-evident but its complete and final resolution ought not to interfere with serious engagement with abolitionist analysis . . .”).

25. Cf. Stéphane Mechoulan & Nicolas Sahuguet, *Assessing Racial Disparities in Parole Release*, 44 J. LEGAL STUD. 39, 49, 70 (2015) (finding that parole boards that use actuarial instruments were not racially biased in their decisions to release black and white prisoners).

in assuaging both of these fears.²⁶ RAIs can help differentiate high-risk and low-risk offenders while at the same time constraining the decisions that do so. In apparent recognition of these possibilities, about half the states use RAIs in some fashion at the dispositional stage or as a means of allocating correctional resources.²⁷

Despite their increasing prevalence, however, the full impact of risk algorithms has yet to be either realized or adequately assessed. That is because their use is, in every jurisdiction, entirely discretionary.²⁸ In Virginia, for instance, a 2016 study found that RAI-based sentencing recommendations were overridden in 40 percent of the cases; judges imprisoned 42 percent of those recommended for alternatives to prison and permitted alternative dispositions for 23 percent of those recommended for imprisonment.²⁹ Other researchers report similar dynamics.³⁰

From a decarceration perspective, that is unfortunate. Researchers with bipartisan credentials who audited the compositions of the prison populations in three states estimated that, if danger to the community were the only justification for continued confinement, roughly half the prisoners would be released.³¹ The

26. For an overview of risk assessment instruments, see HANDBOOK OF VIOLENCE RISK ASSESSMENT (Randy K. Otto & Kevin S. Douglas eds., 2d ed. 2020).

27. John Monahan & Jennifer L. Skeem, *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning*, 26 FED. SENT'G REP. 158, 159–60 (2014); JENNIFER ELEK ET AL., USING RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: OBSERVATIONS FROM TEN JURISDICTIONS 8–9, 17–18, 31–38 (2015).

28. Generally, the statutes say that state officials should “consider” the risk assessment results. See, e.g., KY. REV. STAT. § 532.007(3)(a) (2011) (“Sentencing judges shall consider . . . the result[] of a defendant’s risk and needs assessment included in the presentence investigation”); OHIO REV. CODE § 5120.114 (2019) (“The department . . . shall select a single validated risk assessment tool for adult offenders” to be used by judges at sentencing, corrections officials at the state and local levels, and parole officials); WASH. REV. CODE § 9.94A.500 (2013) (“[T]he [sentencing] court may order the department [of corrections] to complete a risk assessment report,” which “[t]he court shall consider” before imposing a sentence).

29. See Brandon L. Garrett et al., *Judicial Reliance on Risk Assessment in Sentencing Drug and Property Offenders: A Test of the Treatment Resource Hypothesis*, 46 CRIM. JUST. & BEHAV. 799, 807 (2019).

30. Sarah Brayne & Angèle Christin, *Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts*, SOC. PROBS. 1, 11 (2020) (“[M]ost judges and prosecutors did not typically rely on the risk scores at their disposal.”).

31. ANNE MORRISON PIEHL ET AL., CTR. FOR CIVIC INNOVATION, RIGHT-SIZING JUSTICE: A COST-BENEFIT ANALYSIS OF IMPRISONMENT IN THREE STATES 12 (1999); see also BERT USEEM & ANNE MORRISON PIEHL, PRISON STATE: THE CHALLENGE OF MASS INCARCERATION 67 (2008) (concluding, based on three economic cost-benefit studies of five state prison systems, that “[a]lthough it clearly pays on incapacitation grounds alone to incarcerate those at the eightieth percentile [of

Prisoner Assessment Tool Targeting Estimated Risk and Needs (“PATTERN”), the RAI developed in connection with the federal First Step Act, adopts a very narrow definition of low risk, but it nonetheless initially identified 48 percent of its sample population of prisoners as “low” or “minimum” risk and thus eligible for early release from their prison sentences.³² In Virginia, RAIs designed to recommend to judges who should be sentenced to prison alternatives identified 63 percent of drug offenders and 43 percent of larceny and fraud offenders as being low risk, with low risk defined as a 10–15 percent chance of recidivism within the next three years.³³

When offender risk and base rates for reoffending are not quantified in this way, the tendency, at least in the United States, is to opt for incarceration. In particular, judges and prosecutors who are subject to election (which describes almost all state court judges and prosecutors) and parole board members who owe their livelihoods to fickle politicians have good reason to avoid appearing “soft” on crime, given American cultural proclivities.³⁴ As David Ball’s parsing of the psychology literature shows, there is also “the human tendency to desire certainty and simplicity;” Ball suggests that this desire, which he attributes to judges as well as everyone else, “may help explain why [our] default seems to be to keep someone locked up, ‘just in case’—and why this desire is resistant to information and argument.”³⁵

In contrast, the quantified results of well-validated RAIs can provide a concrete, rational basis for diversion or release. If, as recommended in this Article, adherence to those results is *required* in most circumstances, the human urge to incapacitate those in the law’s grasp can be even more effectively resisted because decision-makers must obey the objective facts. Evidence of such a dynamic comes from Virginia, which found that judges using risk algorithms were willing to reduce sentences even for sex offenders “when they can point to the

risk] in all five states, it does not appear to ‘pay’ to incarcerate those below the median”).

32. U.S. DEP’T OF JUST., THE FIRST STEP ACT OF 2018: RISK AND NEEDS ASSESSMENT SYSTEM 58 (2019), <https://www.bop.gov/inmates/fsa/docs/the-first-step-act-of-2018-risk-and-needs-assessment-system.pdf>.

33. VA. CRIM. SENT’G COMM’N, 2012 ANNUAL REPORT 47 (2012), www.vcsc.virginia.gov/2012VCSCAnnualReport.pdf.

34. See Christopher Slobogin, *How Changes in American Culture Triggered Hyper-Incarceration: Variations on the Tazian View*, 58 HOW. L.J. 305, 320–21 (2015); Jeffrey J. Rachlinski & Andrew J. Wistrich, *Judging the Judiciary by the Numbers: Empirical Research on Judges*, 13 ANN. REV. L. & SOC. SCI. 203, 210–11 (2017).

35. W. David Ball, *The Peter Parker Problem*, 95 N.Y.U. L. REV. 879, 879 (2020).

low risk-assessment score as a second opinion to support their decision.”³⁶

For the same reasons, racial and other types of bias in decision-making about postconviction release can be significantly reduced if the relevant cut-points have the force of law. Even if, as some claim, RAIs are more likely to misclassify black people than white people as high risk (a claim that this Article looks at closely), large numbers of black people will still be classified as low risk. If that categorization creates a presumption against incarceration, more people of color will be eligible for release.³⁷ In contrast, a regime that is based on intuitive or clinical judgments about who is “dangerous” is too easily manipulated and prone to overly conservative outcomes influenced by conscious or unconscious prejudices. Indeed, the available research indicates that racial disparity becomes pronounced when, contrary to the recommendation of this Article, legal decision-makers depart from the risk algorithm.³⁸

RAIs can also help identify ways of ameliorating the risk of offenders, whether they are released or confined. The old mantra that “nothing works” in the battle against recidivism has been soundly debunked.³⁹ As one 2008 review of the research summarized it: “The global question of whether rehabilitation treatment works to reduce recidivism has been answered in the affirmative by every meta-analyst who has conducted a systematic synthesis of a broad sample of the available experimental and quasi-experimental research.”⁴⁰ Ten years later, another meta-analysis of institutional programs aimed at reducing the risk of adults confirmed that the recidivism of

36. MEGAN T. STEVENSON & JENNIFER L. DOLEAC, IZA INST. OF LAB. ECON., Ser. No. 12853, ALGORITHMIC RISK ASSESSMENT IN THE HANDS OF HUMANS 20 (2019), <http://ftp.iza.org/dp12853.pdf>; see also Sarah Desmarais & Samantha A. Zottola, *Violent Risk Assessment: Current Status and Contemporary Issues*, 103 MARQUETTE L. REV. 793, 804–07 (2020) (reporting research indicating that the majority of studies show that RAIs, when implemented properly, reduce both use of restrictive conditions and violent recidivism).

37. Kevin R. Reitz, *The Compelling Case for Low-Violence-Risk Preclusion in American Prison Policy*, 38 BEHAV. SCI. & L. 207, 207 (2020).

38. Ben Green & Yiling Chen, *Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments* 3 (2019) (unpublished manuscript), <https://scholar.harvard.edu/files/19-fat.pdf> [<https://perma.cc/Q8QA-AHHL>]; Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 308–09 (2018); Jon Kleinberg et al., *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 241 (2018).

39. See Robert Martinson, *New Findings, New Views: A Note of Caution Regarding Sentencing Reform*, 7 HOFSTRA L. REV. 243, 253–54 (1979) (repudiating the author’s earlier review of research, which had come to stand for the proposition that “nothing works” in terms of reducing recidivism).

40. Mark W. Lipsey & Francis T. Cullen, *The Effectiveness of Correctional Rehabilitation: A Review of Systematic Reviews*, 3 ANN. REV. L. & SOC. SCI. 297, 306 (2007).

adults was roughly 35 percent lower than those who are not treated.⁴¹ In many types of cases, community-based programs are even more effective at curbing violent behavior.⁴² RAIs that identify criminogenic needs can link individuals to the appropriate programs.⁴³

In theory, then, RAIs could bring significant benefits: First, fewer people—of all ethnicities—would be put in jail prior to trial and in prison after conviction. Second, for a substantial number of those who are imprisoned, overall sentences would be shorter. Third, treatment resources would be more efficiently allocated. Fourth, for these three reasons, the capital outlays for prisons and jails would be substantially less (although alternatives to prison, including good treatment programs, would cost more). Fifth, the public would continue to be protected, and perhaps would be even better protected, from the most dangerous individuals. Sixth, lower-risk individuals would be less subject to the well-documented criminogenic effects of incarceration and better positioned to build and maintain a life outside of jail or prison that does not involve criminal activity.

RAIs cannot fully realize these benefits, however, unless the currently popular determinate sentencing structure that exists in most states is dramatically altered. Today, determinate sentencing states give almost all sentencing power to prosecutors, who in essence fix the sentence range through charging practices, and judges, who decide where within the range the sentence will fall and occasionally select a sentence outside that range.⁴⁴ Even in states that technically retain parole, the power of parole boards to affect sentences is very circumscribed.⁴⁵ The result is that even an offender who poses a low risk of reoffending will often receive a lengthy prison sentence.

41. Nina Papalia et al., *A Meta-Analytic Review of the Efficacy of Psychological Treatments for Violent Offenders in Correctional and Forensic Mental Health Settings*, CLINICAL PSYCH.: SCI. & PRAC., June 2019, at 15, <https://onlinelibrary.wiley.com/doi/epdf/10.1111/cpsp.12282>.

42. BEYOND RECIDIVISM: NEW APPROACHES TO RESEARCH ON PRISONER REENTRY AND REINTEGRATION 29 (Andrea Levenretz et al. eds., 2020). For a description of several successful programs, and relevant data, see SHARON MIHALIC ET AL., UNIV. COLO. BOULDER, CTR. FOR THE STUDY AND PREVENTION OF VIOLENCE, BLUEPRINTS FOR VIOLENCE PREVENTION 15–42, 56–58 (2004), <https://www.ncjrs.gov/pdffiles1/ojdp/204274.pdf>.

43. D.A. ANDREWS & JAMES BONTA, THE PSYCHOLOGY OF CRIMINAL CONDUCT 242–43 (6th ed. 2017) (noting that most of these studies were carried out by the developer of the treatment and that replications produced recidivism reduction rates closer to 15 percent); BEYOND RECIDIVISM: NEW APPROACHES TO RESEARCH ON PRISONER REENTRY AND REINTEGRATION, *supra* note 42, at 26–27.

44. *Blakely v. Washington*, 542 U.S. 296, 331 (2004) (Breyer, J., dissenting) (“[I]n a world of statutorily fixed mandatory sentences for many crimes, determinate sentencing gives tremendous power to prosecutors to manipulate sentences through their choice of charges.”).

45. *See infra* text accompanying note 48.

Preventive justice—again, a sentencing scheme that limits the prosecutor and judge to a determination of a sentence range, with the ultimate sentence determined through risk assessments—would restore power to parole boards. The indeterminate sentencing regimes of yesteryear, featuring broad sentence ranges and release decision-making by politically appointed parole board members, were much maligned, justifiably so in many respects. Critics voiced concerns about the incompetence of parole officials and the inaccuracy of predictions, among other objections.⁴⁶ But if instead, parole boards were composed of experts in risk assessment, used RAIs in their assessments, and were required to release all but high risk offenders at the end of the minimum sentence demanded by desert, a much different type of indeterminate sentencing would exist—one that would dramatically reduce the amount of time offenders spend in prison, if they go to prison at all.

Even less refined parole schemes have produced such results. Kevin Reitz has made the argument that much of our prison growth since the 1990s has been in states with indeterminate systems controlled by parole boards, which increasingly refused to release prisoners who were eligible for parole.⁴⁷ But he does not mention the fact that the same pressures that led to truth-in-sentencing, mandatory minima, and three-strikes laws also made “indeterminate” sentencing regimes much more determinate; legislative initiatives in these jurisdictions significantly increased the proportion of prisoners who are ineligible for parole, to anywhere from 45 to 93 percent.⁴⁸ And while parole boards are certainly vulnerable to political pressures from governors and legislatures, that malleability works both ways; Reitz’s work acknowledges that before the tough-on-crime movement in the 1980s and 1990s, parole boards were much more willing to release prisoners early,⁴⁹ and that today

46. Edward E. Rhine, *The Present Status and Future Prospects of Parole Boards and Parole Supervision*, in THE OXFORD HANDBOOK OF SENTENCING AND CORRECTIONS 627, 630–32 (Joan Petersilia & Kevin R. Reitz eds., 2012) (detailing criticism of parole boards, including that parole is “an inherently flawed concept that should be abolished altogether”).

47. Kevin R. Reitz, *Don’t Blame Determinacy: U.S. Incarceration Growth Has Been Driven by Other Forces*, 84 TEX. L. REV. 1787, 1795 (2006).

48. For instance, Reitz says that the top four “carceral powerhouses” were “Texas, Louisiana, Oklahoma and Georgia, all indeterminate jurisdictions.” *Id.* But all of these states significantly limit the proportion of the prison population that is eligible for parole: Texas (55 percent), Louisiana (7 percent), Oklahoma (17 percent) and Georgia (39 percent), and none of these states recognizes presumptive parole (which requires automatic release if certain criteria are met). See JORGE RENAUD, PRISON POL’Y INITIATIVE, FAILURE SHOULD NOT BE AN OPTION: GRADING THE PAROLE RELEASE SYSTEMS OF ALL 50 STATES app. A (2019), https://www.prisonpolicy.org/reports/parole_grades_table.html.

49. Kevin R. Reitz, *Prison-Release Reform and American Decarceration*, 104 MINN. L. REV. 2741, 2745 (2020) (“During the thirty-five-year buildup

the states that are leading the charge in the decarceration movement are those with indeterminate sentencing.⁵⁰ Expert parole board decision-making constrained by RAIs should be able to do even better.

Part II of the Article provides more detail on this preventive justice regime and illustrates how it both resembles and differs from the indeterminate sentencing systems of yore. Parts III and IV grapple with the controversies surrounding RAIs; although some attention will be paid to complaints about the accuracy of these tools, the bulk of the discussion will be about algorithmic fairness, which has been the focus of most criticisms of RAIs. Part V calls on states to experiment with preventive justice approaches. The hypothesis of this Article, which needs to be given a fair test, is that a system of preventive justice offers the single most potent, and most realistic, mechanism for bringing about significant reform of the American criminal punishment system.

II. PREVENTIVE JUSTICE AND LIMITING RETRIBUTIVISM

The term preventive justice, as used in this Article, refers to a sentencing regime that adopts sentence ranges consistent with the

period, . . . parole boards . . . became progressively more risk-averse in their decision-making and ever more fearful of external scrutiny and condemnation. Instead of using their release discretion as often as they had done in the earlier twentieth century, parole boards transformed themselves into agencies of ‘release-denial discretion.’”). This is also why attempts to compare the effects of ‘determinate’ and ‘indeterminate’ sentencing regimes are so conflicted. Compare Don Stemen & Andres F. Rengifo, *Policies and Imprisonment: The Impact of Structured Sentencing and Determinate Sentencing on State Incarceration Rates, 1978-2004*, 28 JUST. Q. 174, 181–82 (2011) (reasoning that studies often find conflicting results regarding these two regimes because they incorrectly focus on the policy motivations rather than the policy structures), with Yan Zhang et al., *Indeterminate and Determinate Sentencing Models: A State-Specific Analysis of Their Effects on Recidivism*, 60 CRIME & DELINQ. 693, 711 (2009) (obtaining mixed results and concluding that “the effects of different sentencing models on reoffending may be largely contingent on . . . supervision approaches . . . differing expertise of state parole boards, or differing crime categories that are legislatively mandated”).

50. Reitz, *supra* note 49, at 2775–76 (“[I]t is intriguing that indeterminate states have had more than twice as much prison-rate decline as determinate states in the post-growth period. This raises the possibility that the low-friction quality of indeterminate prison-release systems could make it easier to reverse course than in stickier determinate regimes.”); Kevin R. Reitz, *Prison-Release Discretion and Prison Population Size State Report: Texas 7* (May 31, 2020) (unpublished manuscript) (on file with author) (“Texas is (undisputably) an example of a jurisdiction with an extremely high degree of indeterminacy overall in its framework for prison-release decision making, even when compared with other paroling states. It is also (undisputably) an example of a state that has dramatically reduced its imprisonment rate over the past 20 years in a nearly continuous downward trend. Even without solid evidence of a causal linkage, this is a brightly-colored flag that further inquiry is warranted.”).

offender's desert and then relies on expert parole boards to determine the nature and duration of sentence within this range, based on consideration of individual prevention goals (i.e., incapacitation, specific deterrence, and rehabilitation) as measured through RAIs.⁵¹ Such a system can alleviate many of the inherent tensions between desert and prevention, between deontology and political reality, and between the desire for community input and the allure of expertise. If done properly, it should also significantly reduce prison populations.

Preventive justice, so defined, is a form of what is often called limiting retributivism. That concept is usually associated with Norval Morris, who almost a half century ago wrote *The Future of Imprisonment*.⁵² In that book, Morris proposed that retributivism should play a “negative” role at sentencing.⁵³ In other words, a person's desert or blameworthiness should act as a constraint on sentencing, but not as a determinant of a particular sentence. As a practical matter, this meant that desert would determine the range of the sentence for particular crimes, but not the disposition in a particular case. Rather, Morris argued, the specific sentence should depend on other, utilitarian considerations—although, importantly, in contrast to the sentencing schemes popular at the time and to preventive justice, that sentence was to be set at the front end, not determined by a parole board at the back end.⁵⁴ Sentences would also be guided by what Morris called the “parsimony principle”—the idea, endorsed both before and since by many commentators—that punishment should be no more severe than necessary to carry out the state's punishment purposes.⁵⁵ Morris's limiting retributivism has often been called a “hybrid” approach that attempts to reconcile, or at least combine, deontologically driven desert theory with the

51. I coined the phrase “preventive justice” in 2011 to refer to the type of sentencing described in the text. See CHRISTOPHER SLOBOGIN & MARK R. FONDACARO, *JUVENILES AT RISK: A PLEA FOR PREVENTIVE JUSTICE* 6–7, 63–64 (2011). Other authors have used it in a pejorative sense, primarily as a synonym for preventive detention. See ANDREW ASHWORTH & LUCIA ZEDNER, *PREVENTIVE JUSTICE* 144–45 (2014) (using the term as a synonym for preventative detention); Laurence H. Tribe, *An Ounce of Detention: Preventive Justice in the World of John Mitchell*, 56 VA. L. REV. 371, 392 (1970).

52. See NORVAL MORRIS, *THE FUTURE OF IMPRISONMENT* 75 (1974) (theorizing that inflicted punishment should be no more severe than necessary and that, even then, it is not obligatory to punish a criminal if utilitarian factors dictate otherwise).

53. *Id.* at 73.

54. *Id.* at 47–49, 75.

55. *Id.* at 60–62; see also JEREMY BENTHAM, *THEORY OF LEGISLATION* 345 (R. Hildreth trans., London, Trübner & Co. 5th ed. 1887) (1802); RICHARD S. FRASE, *JUST SENTENCING: PRINCIPLES AND PROCEDURES FOR A WORKABLE SYSTEM* 32–33 (2012).

consequentialist goals of protecting the public and conserving public resources.⁵⁶

Morris's version of limiting retributivism, which was a reaction to the indeterminate sentencing regimes that reigned at the time he wrote, proved to be quite popular. A number of states adopted a version of limiting retributivism beginning in the 1970s.⁵⁷ For instance, in 1980, Minnesota created sentencing guidelines consistent with Morris's view.⁵⁸ The state's guidelines have eleven categories of sentences, which provide fairly narrow ranges (about 20 months) for lower level felonies, and fairly broad ranges (about 120 months) for higher level felonies.⁵⁹ The lower end of the most punitive category approximates the higher end of the next most punitive category, the lower end of the second most punitive category abuts the higher end of the next category, and so on. Most lower-level felonies result in imposition of sentences that are then suspended, at least for first time offenders. Numerous states have similar sentencing guidelines, although many adopted ranges that were much smaller than Minnesota's.⁶⁰ The 2017 revisions to the sentencing provisions of the Model Penal Code ("MPC") likewise call for numerous sentencing categories defined by proportionate culpability.⁶¹ While the MPC would permit risk assessments to influence the sentence if they are based on "sufficiently reliable" tools, that input would be considered by the judge, whose sentence would be determinate and set at the front end, not subject to change by a parole board.⁶²

Like these other limiting retributivist schemes, the type of preventive justice I propose would have sentencing categories based on proportionate culpability. But it differs in three ways: First, given the nebulousness of the culpability inquiry, the ranges would be fewer and broader than in many versions of limiting retributivism. Second, sentence length within the range would be determined by a parole board at the back end, not by a judge at the front end. Third, the focus of the parole board would be solely on risk and needs assessment. Retributive (and general deterrence) concerns would be solely legislative matters, integrated into the statutes governing

56. FRASE, *supra* note 55, at 82–85.

57. Richard S. Frase, *Sentencing Guidelines in the States: Lessons for State and Federal Reformers*, 6 FED. SENT'G REP. 123, 123 (1993) (noting that seventeen states have adopted "presumptive sentencing rules").

58. MINN. SENT'G GUIDELINES COMM'N, MINNESOTA SENTENCING GUIDELINES AND COMMENTARY 7 (2019), <https://mn.gov/msgc-stat/documents/Guidelines/2019/MinnSentencingGuidelinesCommentary.pdf>.

59. *Id.* at 79.

60. See Richard S. Frase, *Sentencing Guidelines in Minnesota, 1978–2003*, 32 CRIME & JUST. 131, 206 (2005).

61. MODEL PENAL CODE: SENTENCING §§ 1.02(2), 6.01, 6.11 (AM. L. INST. 2017).

62. *Id.* §§ 6.11, 9.08.

sentence ranges. They would not inform sentences in individual cases.

The model for preventive justice comes from the sentencing provisions in the *original* Model Penal Code, which was promulgated in 1962. The sentencing provisions in that version of the MPC mandated only three broad sentencing ranges—for first-degree felonies, one year up to life in prison (or up to twenty years, depending on which version of the MPC was adopted); one to ten years for second-degree felonies; and one to five years for third-degree felonies—and in each case the one-year minimum could involve an alternative to prison.⁶³ The parole board determined the ultimate sentence length, based primarily on individual prevention considerations.⁶⁴ The primary difference between the original MPC's approach and the version of preventive justice sketched out here is that, in the latter regime, RAIs would modernize the parole board determination. Another difference is that, to give retributive goals sufficient due, more sentencing categories would exist.

The various differences between preventive justice and other forms of limiting retributivism can be fleshed out by looking more closely at three issues: the determination of sentencing ranges, the principles—constitutional and otherwise—that would govern preventive justice, and the nature of the parole board in a preventive justice regime.

A. *The Determination of Sentence Ranges*

The first goal in any limiting-retributivism regime, whether or not mixed with risk assessment, is to set the sentence range for each type of crime. At one end of the spectrum are scholars like Andrew von Hirsch and Andrew Ashworth, both retributivists, who believe that “parity” in sentencing is crucial. They thus call for very narrow sentence ranges based entirely on “the degree of harmfulness of the conduct and the extent of the actor’s culpability.”⁶⁵ Although they would allow “limited variations . . . , say, in the range of 5–10 percent” to permit non-prison alternatives and “back-up sanctions” for offenders who violate probation or parole conditions, they argue that preventive aims should never be grounds for deviation.⁶⁶ At the other end of the spectrum is Morris, who apparently found acceptable very wide ranges, akin to those associated with indeterminate regimes. At least in his early writings, Morris held to the position that blameworthiness is such a capacious construct that a given crime could be associated with a large variety of sanctions that are, as he

63. MODEL PENAL CODE §§ 6.02, 6.06 (AM. L. INST., Proposed Official Draft 1962).

64. *See id.* §§ 305.6–305.9.

65. *See* ANDREW VON HIRSCH & ANDREW ASHWORTH, PROPORTIONATE SENTENCING: EXPLORING THE PRINCIPLES 144, 159 (2005).

66. *Id.* at 161–62.

put it, “not undeserved.”⁶⁷ In-between are writers like Richard Frase, who endorses something akin to Minnesota’s system of moderately broad ranges. Frase thinks that Morris was too casual in his attitude toward desert, but he also rejects the narrow ranges proposed by von Hirsch and Ashworth, not only because he is willing to differentiate between high and low risk offenders but also because he believes broader ranges are necessary to encourage offender cooperation and to provide leeway in meeting correctional priorities.⁶⁸

The wide ranges preferred by Morris make the most sense. Any other approach ignores the hollowness of assigning specific punishments to specific crimes. Von Hirsch and Ashworth admit as much, conceding that “[t]here seems to be no crime for which one can readily perceive a specific quantum of punishment as the uniquely deserved one.”⁶⁹ Nonetheless, they claim that precision is still possible because consensus can at least be reached about the “ordinal” ranking of crimes.⁷⁰ Thus, once the maximum “anchor” point (for instance, life without parole) has been set, punishments can be assigned according to their ranking, with the anchor point providing the upper limit on punishment for the most serious crime and lesser punishments assigned to lesser crimes.⁷¹ The problem with this approach is that consensus on desert is simply not possible. While there is widespread agreement with respect to the ordinal rankings of the most common crimes (such as homicide, rape, robbery, and theft), outside of those core crimes, consensus over ordinal rankings disappears.⁷² More importantly, disagreement over the all-important maximum anchor point can be significant; for instance, some might favor the death penalty for murder, while others might prefer a maximum sentence of fifteen years for such a crime. Finally, the “spacing” between crimes on the ordinal scale can also be a matter of serious dispute.⁷³

67. NORVAL MORRIS, MADNESS AND THE CRIMINAL LAW 150–52 (1982).

68. FRASE, *supra* note 55, at 29–30.

69. VON HIRSCH & ASHWORTH, *supra* note 65, at 142.

70. *Id.* at 141–43.

71. *Id.* at 142.

72. See Paul H. Robinson & Robert Kurzban, *Concordance and Conflict in Intuitions of Justice*, 91 MINN. L. REV. 1829, 1890–92 (2007).

73. In a study I conducted with Lauren Brinkley-Rubinstein, subjects were asked to look at twelve crime scenarios describing only the *actus reus* and *mens rea* for the crime, and then to indicate appropriate punishment on a thirteen-point scale. Christopher Slobogin & Lauren Brinkley-Rubinstein, *Putting Desert in its Place*, 65 STAN. L. REV. 77, 94–96 (2013). Not surprisingly, agreement on the proper punishment was virtually nonexistent. See *id.* at 95 tbl.2 (reporting that for the “Control Group,” which received scenarios describing the *actus reus* and *mens rea*, standard deviations ranging from 3.109 to 7.373). And the range of sentences was also broad, even when the extreme dispositions beyond two standard deviations were thrown out. See *id.* (reporting, for example, ranges from zero to four years for theft of a pie and from one day to the death penalty for

Von Hirsch and Ashworth suggest that the ordinal ranking and spacing questions can be addressed through analyzing the impact of a given crime on the victim's "standard of living," a concept they develop at some length.⁷⁴ But, as they acknowledge, "the impact of a crime on the living standard is itself very much a matter of factual and normative judgement" and in any event "is designed mainly to address crimes directly affecting natural persons [e.g., homicide, sex, and property crimes], rather than crimes involving collective interests."⁷⁵ More importantly, the standard of living metric addresses only the *actus reus* of the crime. Von Hirsch and Ashworth suggest that the mental state component, which can be an extremely important factor in determining desert for many crimes, instead "can be gauged with the aid of clues from substantive criminal law doctrine."⁷⁶ But, again, the parameters of an offender's blameworthiness are highly debatable, and the ability to discern and correctly label the relevant facts is far from perfect.

While these difficulties do not mean that we should not try to reach conclusions about these matters, they do suggest that any answers attained have a wide margin of error, which should be reflected in sentencing ranges, akin to those found in the original MPC. However, a bit more nuance is advisable if retributivist instincts are to be given adequate weight. First, recall that minimum felony sentences in the original MPC were all one year, regardless of how serious the crime was, and that even a one-year sentence could be served outside of prison.⁷⁷ That outcome is too flexible if one takes ordinal desert and the goal of differentiating the worst from the least crimes seriously, and it is drastic enough that it could significantly undermine the general deterrent impact of specific sentences, to the extent that there is any.⁷⁸ Additionally, as Paul Robinson argues from a consequentialist perspective, a punishment system that departs too far from community sentiment may well lose legitimacy and produce a populace more prone to take the law into its own hands or to become noncompliant in other ways.⁷⁹ Although I have argued

a killing by pit bulls whose owner allowed them to escape). Further, in only the least and most serious crime scenarios did more than 25 percent of the sample choose the same punishment. *See id.*

74. VON HIRSCH & ASHWORTH, *supra* note 65, at 143–46.

75. *Id.* at 146.

76. *Id.* at 144.

77. MODEL PENAL CODE §§ 6.02, 6.06 (AM. L. INST., Proposed Official Draft 1962).

78. For a general critique of the conceptual and empirical basis of general deterrence as a basis for determining criminal punishment, see Paul H. Robinson & John M. Darley, *The Role of Deterrence in the Formulation of Criminal Law Rules: At Its Worst When Doing Its Best*, 91 GEO. L.J. 949, 950–51 (2003).

79. *See* Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. L. REV. 453, 498 (1997) ("We conclude that desert distribution of liability happens to be the distribution that has the greatest utility, in the sense of avoiding crime.

that Robinson overstates the need for adhering strictly to societal views on punishment,⁸⁰ a default sentence of one year for a crime like aggravated murder might well trigger the type of delegitimization he describes.⁸¹

Taking these considerations into account, a limiting retributivist regime might adopt ranges for felonies like the following: ten to thirty years for very serious crimes (such as, and perhaps only for, aggravated murder), followed by four additional tiers of felonies, with ranges of three to ten years (for lesser forms of intentional killing, rape, and armed robbery), one to five years (for crimes like aggravated assault), one to three years (for crimes like burglary) and up to one year (for, e.g., the typical low-level drug sale crime). The minima and maxima might be discounted for youth or for other concrete mitigating circumstances (such as providing substantial assistance to the authorities). They might also be extended for offenders who have committed multiple crimes or for other concrete aggravating circumstances, *if* that extension is justified solely on blameworthiness grounds (with respect to risk, previous crimes should be given effect only through an RAI).⁸² Further, the relevant time periods need not be served in prison; perhaps for all but the first category, alternatives to prison should be an option. Undoubtedly, many readers will quibble (or perhaps more than quibble) over these ranges and examples, which should bring home the fact that desert is highly malleable; obviously, legislatures or sentencing commissions that find these choices either too harsh or too lenient could modify them accordingly.

While this type of adjustment to the minima and the addition of two categories to the MPC's original three should alleviate the most significant concerns about disproportionality, these modifications would not address von Hirsch and Ashworth's concern about disparate sentences. But, again, this objection assumes both that the

Thus, utility theorists ought to support liabilities assigned according to such a desert-based system"); Paul H. Robinson et al., *The Disutility of Injustice*, 85 N.Y.U. L. REV. 1940, 1948 (2010) (same).

80. Slobogin & Brinkley-Rubinstein, *supra* note 73, at 96–110 (describing research debunking Robinson's assertion showing lack-of-compliance effects of an empirical desert approach).

81. *Id.* at 118 (concluding, based on empirical findings, that "[w]here serious crimes such as murder are involved, desert appears to play a much more dominant role for a majority of people"); see also Dena M. Gromet & John M. Darley, *Restoration and Retribution: How Including Retributive Components Affects the Acceptability of Restorative Justice Procedures*, 19 SOC. JUST. RSCH. 395, 399 (2006) (finding that as crimes increase in seriousness, people are more likely to require a retributive component).

82. On whether sentence enhancements based on prior crimes can be justified on retributive grounds, see Julian V. Roberts & Orhun H. Yalincak, *Revisiting Prior Record Enhancement Provisions in State Sentencing Guidelines*, 26 FED. SENT'G REP. 177, 177 (2014).

desert of two different offenders can be rigorously measured and that desert is the relevant metric for gauging parity. If instead it is taken as a given that calibrating desert for a specific crime is an abstruse exercise (consider, for instance, the many possible variations of “armed robbery” and the people who commit it) and that the whole point of limiting retributivism is to recognize that other criminal justice considerations besides desert may legitimately be considered in fashioning sentences, this concern is appreciably diminished.

That is not to say that parity is not an important goal. For instance, as developed below, offenders with similar risk levels should be treated similarly, regardless of other characteristics. More generally, the existence of wide sentence ranges and specific sentences based on nondesert considerations should not lead to the conclusion that sentencing can be “lawless,” the term Judge Frankel famously affixed to indeterminate regimes.⁸³ Several principles, some of them arguably of constitutional status, should govern a preventive justice version of limiting retributivism.

B. Governing Principles of Preventive Justice

The first principle of a preventive justice version of limiting retributivism involves the desert component rather than the prevention component, and it would be straightforward: to satisfy the retributive goal, the minimum sentence required by desert must be served. For most misdemeanors and many lower level felonies (those in the final three of the five categories proposed above), this sentence might often consist of nonincarcerative restrictions in the community, such as community service and restitution; here, attention should be paid to the fact that some types of community restrictions can have as much “punitive bite” as short prison sentences.⁸⁴ Frase argues, as did Henry Hart before him,⁸⁵ that the censure associated with a criminal conviction and these types of dispositions can satisfy desert demands at the lower end of the criminal scale.⁸⁶ Several other commentators, including von Hirsch and Ashworth, have recognized

83. MARVIN E. FRANKEL, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* 3–11 (1973) (criticizing indeterminate sentencing and stating, *inter alia*, that “legislatures have not done the most rudimentary job of enacting meaningful sentencing ‘laws’,” and that, as a result, sentencing judges and parole officials exercised broad discretion and wielded enormous sentencing power “effectively subject to no law at all”).

84. Robert E. Harlow et al., *The Severity of Intermediate Penal Sanctions: A Psychophysical Scaling Approach for Obtaining Community Perceptions*, 11 J. QUANTITATIVE CRIMINOLOGY 71, 87 (1995).

85. Henry M. Hart, Jr., *The Aims of the Criminal Law*, 23 L. & CONTEMP. PROBS. 401, 427 (1958).

86. FRASE, *supra* note 55, at 31.

that retributive goals can be achieved through nonincarcerative alternatives.⁸⁷

Furthermore, once granted, probation or parole should be revoked only if an individual's risk category changes. In some jurisdictions, almost half of all prison admissions are for parole violations that can consist of very minor violations having nothing to do with the degree of risk that ought to be the basis for detention.⁸⁸ In a preventive justice regime, those types of detentions would be impermissible.

However, for more serious crimes (certainly those in the first category and perhaps most of those in the second), probably no community-based condition would satisfy desert, in which case the presumptive sentence would be incarceration for the minimum period of the relevant sentencing range. Frase argues in favor of the practice in many states of making the default sentence the middle of the range, so that judges can depart upward or downward and still stay within desert-based bounds.⁸⁹ But if, as we are assuming, any sentence within the range satisfies desert, the parsimony principle—which writers like Morris, Tonry, and Frase himself all endorse (as does the revised MPC)⁹⁰—dictates that the minimum sentence is the place to start. Building on the idea that culpability is a hopelessly vague concept, Jacob Bronsther has also concluded “that retributivism can only justify the *least harmful* ‘not undeserved’ sentence.”⁹¹

The parsimony principle, reinforced by constitutional doctrine, would also govern the role of risk assessment and risk management. Almost fifty years ago, the Supreme Court decided *Jackson v.*

87. VON HIRSCH & ASHWORTH, *supra* note 65, at 156–59.

88. See HUMAN RIGHTS WATCH, REVOKED: HOW PROBATION AND PAROLE FEED MASS INCARCERATION IN THE UNITED STATES (2020), <https://www.hrw.org/report/2020/07/31/revoked/how-probation-and-parole-feed-mass-incarceration-united-states> (finding, in a study of three states, that this is the case in Pennsylvania and further noting that “most people locked up for supervision violations were not convicted of new offenses—rather, they were incarcerated for breaking the rules of their supervision, such as for using drugs or alcohol, failing to report address changes, or not following the rules of supervision-mandated programs. Of those who were incarcerated for new offenses, in our focus states, many were for conduct like possessing drugs; public order offenses such as disorderly conduct or resisting arrest; misdemeanor assaultive conduct; or shoplifting.”). On the degree of risk required, see *infra* text accompanying note 260.

89. FRASE, *supra* note 55, at 52.

90. MODEL PENAL CODE: SENTENCING § 1.02(2) (AM. L. INST., Proposed Final Draft 2017) (providing that sentences may be “no more severe than necessary”).

91. Jacob Bronsther, *Vague Comparisons and Proportional Sentencing*, 25 LEGAL THEORY 26, 26 (2019). Also known as Ockam's Razor, the parsimony principle emphasizes that the simplest solution is often the correct one. Elliott Sober, *The Principle of Parsimony*, 32 BRIT. J. PHIL. SCI. 145, 145–46 (1981).

Indiana,⁹² involving a successful challenge to the prolonged detention of an individual who had been found permanently incompetent to stand trial.⁹³ In concluding that Jackson's detention was unconstitutional, the Court declared that the Due Process Clause requires that the "nature and duration of commitment bear some reasonable relation[ship] to [its] purpose."⁹⁴ This pronouncement, repeated in several subsequent Supreme Court decisions,⁹⁵ has at least three implications for any sentence based on a prevention rationale of the type at issue here.

First, *Jackson* can be read to require, consistent with the parsimony principle, that any government deprivation of liberty in the name of prevention be the last drastic means of achieving the state's preventive aim.⁹⁶ That would mean, for instance, that once the desert minimum has been met, there should be a presumption of release, rebuttable only by proof that the person poses a significant risk of reoffending—a proposition some lower courts have adopted.⁹⁷ Similarly, even if the individual is incarcerated and considered high risk when the minimum expires, community-based programs should be the presumptive disposition unless they cannot adequately protect the public. While imprisonment might substantially reduce risk, it may not be any more effective at doing so than placement in a substance abuse treatment program in the community, a vocational training program in a halfway house, or a job-release program coupled with an ankle monitor. *Jackson's* parsimony principle requires consideration of such options.⁹⁸

92. 406 U.S. 715 (1972).

93. *Id.* at 717–19.

94. *Id.* at 738.

95. *Seling v. Young*, 531 U.S. 250, 265 (2001) (citing *Jackson* and indicating that residents of a commitment facility may have a claim for release if they are not receiving treatment); *Foucha v. Louisiana*, 504 U.S. 71, 78–79 (1992) (citing *Jackson*, and holding that a person acquitted by reason of insanity may not be committed as dangerous without a showing of mental illness); *Youngberg v. Romeo*, 457 U.S. 307, 324 (1982) (citing *Jackson* and holding that "the State is under a duty to provide respondent with such training as an appropriate professional would consider reasonable to ensure his safety and to facilitate his ability to function free from bodily restraints").

96. Christopher Slobogin, *Prevention as the Primary Goal of Sentencing: The Modern Case for Indeterminate Dispositions in Criminal Cases*, 48 SAN DIEGO L. REV. 1127, 1138–39 (2011) (arguing that, if the government's objective is to prevent harm, "confinement may occur only if necessary to achieve prevention of harm and may continue only if it remains necessary to achieve that aim").

97. *See Carrillo v. Fabian*, 701 N.W.2d 763, 771–73 (Minn. 2005) (holding that postponing release beyond the initial parole determination is "a significant departure from the basic conditions of the inmate's sentence" and that "under the Due Process Clause of the United States Constitution, [a prisoner] has a protected liberty interest in his [presumptive] release date that triggers a right to procedural due process before that date can be extended").

98. *Jackson*, 406 U.S. at 738.

Second, *Jackson* means that the government must provide rehabilitative services if they will reduce the length of the intervention, because otherwise the confinement is not the least restrictive way of accomplishing the state's aim.⁹⁹ In other words, whether or not a state provides such services during the phase of punishment that is designed to meet the desert-based minimum, *Jackson* mandates a right to ameliorative treatment if a sentence is prolonged past the minimum because of a risk assessment. It is unlikely that this principle could be transformed into an affirmative duty on the part of the state to create treatment services.¹⁰⁰ But it could create an incentive to provide at least those services that are available, because the state will know release is the alternative. In *Kansas v. Hendricks*,¹⁰¹ the Supreme Court held that where incapacitation is the goal, treatment need not occur if it is not "available" or "possible," but suggested that if available treatment is not offered, continued preventive detention would violate due process.¹⁰² In *Seling v. Young*,¹⁰³ the Court made the same point.¹⁰⁴ Admittedly, both of these cases involved postsentence "civil" commitment of sex offenders on dangerousness grounds.¹⁰⁵ But in a preventive justice regime, the portion of a sentence that occurs after the minimum is served is based solely on risk and criminogenic needs. If treatment of needs is reasonably possible, it should be provided.

Finally, *Jackson's* reasonable relationship requirement means that any confinement that occurs should be proportionate to the risk, and that the longer the intervention extends, the greater the proof of risk that must be shown (a precept that could be called risk proportionality, analogous to desert proportionality).¹⁰⁶ This principle should mean that individuals considered to be low risk or at risk of committing only minor crimes should never be detained for

99. Eric S. Janus & Wayne A. Logan, *Substantive Due Process and the Involuntary Confinement of Sexually Violent Predators*, 35 CONN. L. REV. 319, 358 (2003) ("[T]he *Jackson* line of cases suggests a strong right to treatment, one in which *effective* treatment facilitates real progress toward community re-entry.").

100. *But see* Thomas S. v. Morrow, 601 F. Supp. 1055, 1059 (W.D.N.C. 1984) (holding that the Constitution was violated when the state provided only any available treatment, rather than "the *appropriate* treatment").

101. 521 U.S. 346 (1997).

102. *Id.* at 367 ("[T]he State has a statutory obligation to provide 'care and treatment for [persons adjudged sexually dangerous] designed to effect recovery.'") (quoting *Allen v. Illinois*, 478 U.S. 364, 369 (1986)).

103. 523 U.S. 250 (2001).

104. *Id.* at 265 (stating that, if the purpose of a civil commitment statute is "to incapacitate and to treat . . . due process requires that the conditions and duration of confinement under the [statute] bear some reasonable relation to the purpose for which persons are committed").

105. *Id.* at 255–58; *Hendricks*, 521 U.S. at 353–56.

106. CHRISTOPHER SLOBOGIN, *MINDING JUSTICE: LAWS THAT DEPRIVE PEOPLE WITH MENTAL DISABILITY OF LIFE AND LIBERTY* 143–50 (2006).

preventive purposes. Even individuals considered to be high risk could not be confined for prolonged periods of time without increasingly higher showings of risk (and again, no one could be detained beyond the retributively determined maximum sentence).¹⁰⁷ Thus, a high risk of serious crime should be required to prolong the sentence for any significant amount of time, and risk would have to be reassessed periodically.

Morris was reticent about allowing risk assessment to influence sentences because of the legitimate concern that assessing a person's dangerousness while in prison may not reflect his or her dangerousness outside of it.¹⁰⁸ But risk assessment done correctly looks at preprison information as well as at participation and completion of programs while in prison, as the PATTERN does.¹⁰⁹ Moreover, individuals who are not high risk or who appear to be manageable in the community despite their high risk can be evaluated outside the prison setting; in particular, conditional release programs can facilitate efforts to assess risk within community settings.¹¹⁰

Unlike desert, risk is not a static variable. Setting a determinate, unchangeable sentence, whether at the front end or at the expiration of the minimum sentence, is inconsistent with a risk-oriented regime. Within the desert-based maximum, the outer boundary of a sentence in a preventive justice regime would be determined by periodic assessments of an offender's risk, risk which, because of the parsimony principle, would need to be increasingly more serious to authorize prolonging the intervention.

C. Parole as a Constitutional Right

From the foregoing, it should be clear that the parole board is crucial in a preventive justice regime. One of the reasons indeterminate sentencing fell into disfavor was the documented tendency of parole boards to act in a biased, arbitrary manner.¹¹¹ As Steven Chanenson put it, "[p]arole release has historically been an unstructured and wildly discretionary power, subject to the same kinds of irrationalities and abuses that afflict old-style, fully discretionary judicial sentencing on the front end."¹¹²

107. See *id.*; Slobogin, *supra* note 96, at 1135–37.

108. MORRIS, *supra* note 52, at 35–36.

109. U.S. DEP'T OF JUST., *supra* note 32, at 44–45.

110. Conditional release programs have long been a mainstay of preventive detention for people found not guilty by reason of insanity. See, e.g., Jeffrey Rogers et al., *After Oregon's Insanity Defense: A Comparison of Conditional Release and Hospitalization*, 5 INT'L J.L. & PSYCHIATRY 391, 392 (1982).

111. See Ghandnoosh, *supra* note 7, at 141.

112. Steven L. Chanenson, *Guidance from Above and Beyond*, 58 STAN. L. REV. 175, 187 (2005); Kevin R. Reitz, *Questioning the Conventional Wisdom of Parole Release Authority*, in THE FUTURE OF IMPRISONMENT 199, 228 (Michael Tonry ed.,

However, if parole boards are conceived of as part of the risk assessment and management team, they would look and act quite differently than the parole boards Chanenson describes. Not only would they rely on modern risk assessment, but they would also be composed primarily of experts in risk-needs assessment, rather than laypeople and correctional officials.¹¹³ Furthermore, rather than the wide-ranging power typically granted such boards, which are usually permitted to consider retributive and general deterrence objectives as well as risk,¹¹⁴ the parole board in a preventive justice regime would focus entirely on individual prevention goals.

A potential obstacle to making the parole board the centerpiece of a sentencing system is the Supreme Court's stance, dating from the 1980s, that there is no constitutionally recognized right to a parole hearing, much less parole release upon specified criteria.¹¹⁵ Nor is there a right to counsel, witnesses, or cross-examination at any hearing that takes place.¹¹⁶ Thus, there is the danger that a state will adopt a preventive regime but fail to provide the primary mechanism for making it work—an expert, formal parole decision-making process.

In the past twenty-five years, however, a number of Supreme Court decisions have provided a basis for changing this situation,

Oxford Univ. Press 2004) (noting that parole boards have a history of poor process and patronage appointments); *see also* Kenneth Culp Davis, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY 126–33 (1976) (calling the federal parole board the most disappointing agency the author had ever encountered).

113. *See* Stefan J. Bing, Note, *Reconsidering State Parole Board Membership Requirements in Light of Model Penal Code Sentencing Revisions*, 100 KY. L.J. 871, 888 (2012) (arguing that, given the movement toward evidence-based sentences, parole boards, should include “social scientists in the fields of sociology, psychology, or statistics”); *see also* Edward E. Rhine, Joan Petersilia & Kevin R. Reitz, *The Future of Parole Release*, 46 CRIME & JUST. 279, 283 (2017) (arguing the same, although contrary to the proposal here, limiting the parole board to a single decision at the end of the minimum sentence, and capping a sentence enhancement to an amount equal to 25–33 percent of the maximum allowable sentence for the offense, the upper end of the relevant range, or fifteen years, whichever comes first).

114. W. David Ball, *Heinous, Atrocious, and Cruel: Apprendi, Indeterminate Sentencing, and the Meaning of Punishment*, 109 COLUM. L. REV. 893, 903–04 (2009) (stating that “[n]one of the due process cases limits the parole board’s authority to its core institutional competence—measuring an offender’s threat to public safety—and parole boards can therefore find facts that sound in both retribution and dangerousness” and finding that, in California, parole boards routinely do so).

115. *Hewitt v. Helms*, 459 U.S. 460, 467–68 (1983) (“[T]here is no ‘constitutional or inherent right’ to parole . . .” (citations omitted)); *Conn. Bd. of Pardons v. Dumschat*, 452 U.S. 458, 464 (1981) (observing that there is no constitutional entitlement to parole).

116. *See Greenholtz v. Inmates of Neb. Penal & Corr. Complex*, 442 U.S. 1, 15 (1979).

both substantively and procedurally. In *Graham v. Florida*,¹¹⁷ the Court held unconstitutional a mandatory life without parole sentence for juveniles who commit homicide,¹¹⁸ and in *Miller v. Alabama*,¹¹⁹ it extended that ruling to bar all mandatory life without parole sentences for juveniles.¹²⁰ The Court's rationale for these decisions turned in large part on its assessment of the ways in which mandatory life sentences are inconsistent with the goals of punishment, two of which—incapacitation and rehabilitation—virtually require parole hearings. As the *Graham* Court explained:

Even if the State's judgment that Graham was incorrigible were later corroborated by prison misbehavior or failure to mature, *the sentence was still disproportionate because that judgment was made at the outset*. A life without parole sentence improperly denies the juvenile offender a chance to demonstrate growth and maturity. Incapacitation cannot override all other considerations, lest the Eighth Amendment's rule against disproportionate sentences be a nullity. . . .

[Nor can a] sentence of life imprisonment without parole . . . be justified by the goal of rehabilitation. The penalty forswears altogether the rehabilitative ideal. By denying the defendant the right to reenter the community, the State makes an irrevocable judgment about that person's value and place in society. This judgment is not appropriate in light of a juvenile nonhomicide offender's capacity for change and limited moral culpability.¹²¹

The Court concluded: “A State need not guarantee the offender eventual release, but if it imposes a sentence of life it must provide him or her with some realistic opportunity to obtain release before the end of that term.”¹²²

Of course, *Graham* and *Miller* made clear they were limiting their reach to mandatory life without parole sentences for juveniles. They were not addressing other types of mandatory sentences nor adult sentencing. But once an offender has served the minimum demanded by retributive and deterrence objectives, the Court's concern about the irreconcilability of irrevocable sentences with the incapacitative and rehabilitative goals of punishment logically affects adult sentences as well. As Richard Bierschbach argues,

Graham recognized the significance of parole—and, implicitly, back-end sentencing generally—to the constitutional regulation

117. 560 U.S. 48 (2010).

118. *Id.* at 82.

119. 567 U.S. 460 (2012).

120. *Id.* at 489.

121. *Graham*, 560 U.S. at 74 (emphasis added).

122. *Id.* at 82.

of punishment. . . . If taken seriously, *Graham's* view of parole could result in more textured, considered, and just sentences, ones that seek not only to deter and condemn but also to further restoration, reconciliation, and other soft but important values the Court saw as being bound up in parole.¹²³

Parole review should also be relatively frequent for those offenders who are not released at the end of the minimum term and for those who are subject to supervision after release. In *Garner v. Jones*,¹²⁴ the Supreme Court held that a prolonged hiatus between parole hearings violates the Ex Post Facto Clause if it “create[s] a significant risk of increased punishment” relative to the sentence contemplated at the time of sentencing.¹²⁵ If a judge sentences an offender to an indefinite term that presumptively ends when the minimum is reached, as proposed here, a failure of the parole board to conduct frequent periodic reviews after that point could violate that rule. That notion is bolstered by the Supreme Court’s decision in *Kansas v. Hendricks*.¹²⁶ In upholding the constitutionality of postsentence commitment of sexually violent offenders—commitments based entirely on the offender’s risk, just as those in a preventive regime would be after the minimum sentence is served—*Hendricks* indicated that “constitutional provisions for care and treatment” require release when the offender is no longer dangerous and noted that the statute in question effectuated that rule by entitling those committed to “*annual* review to determine whether continued detention was warranted.”¹²⁷

Even if all of this is conceded, parole boards still might not provide much in the way of review if the procedures remain as anemic as they have been historically. Assuming the prescriptions advanced later in this Article are followed, most questions about RAIs should be settled jurisdiction-wide, so that case-by-case litigation on generalizable aspects of risk assessment would not be necessary. Nonetheless, procedural injustice is still likely unless offenders are afforded a more robust adversarial process than currently exists, one

123. Richard Bierschbach, *Proportionality and Parole*, 160 U. PA. L. REV. 1745, 1788 (2012); *see also* Flores v. Stanford, 18 CV 2468 (VB), 2019 WL 4572703, at *8 (S.D.N.Y. Sept. 20, 2019) (holding that, assuming that parole board members “‘have denied, and continue to deny, juvenile lifers release to parole supervision based only on the crime committed or juvenile criminal history,’ and ‘despite clear evidence of rehabilitation and maturity,’” a plausible Eighth Amendment violation has been alleged).

124. 529 U.S. 244 (2000).

125. *Id.* at 255 (stating that a prisoner must show that a rule prolonging a time in prison without a parole hearing “created a significant risk of increasing his punishment”).

126. 521 U.S. 346 (1997).

127. *Id.* at 353 (emphasis added).

that meaningfully enables them to challenge inaccurate information or improper adjustments to the RAI.

Once again, Supreme Court caselaw—this time going back almost fifty years—might provide a wedge, given the fact that, in a preventive justice regime, the offender is entitled to release at the completion of the minimum sentence unless the board determines he or she poses a high risk of reoffending. Under these circumstances, the parole determination is more akin to the determination made at a parole *revocation* hearing. In *Morrissey v. Brewer*,¹²⁸ the Court required much more process at such proceedings, including the rights to notice, to be present, to proffer witnesses and documentary evidence, to confront and cross-examine adverse witnesses, and to obtain “a written statement by the factfinder as to the evidence relied on and reasons for revoking parole.”¹²⁹ And in a later case, the Court held that individuals subject to parole and probation revocation proceedings also have a right to counsel when “a disputed issue can fairly be represented only by a trained advocate.”¹³⁰ That will almost always be true at the initial parole hearing when the results of the first postconviction RAI are considered, and in many cases at subsequent hearings as well, especially if the claim is that *Jackson’s* dictates are being violated.¹³¹

The resources for this more formal parole process could come from the decline in appeals of judicially imposed sentences. Such a decline is highly likely in a preventive justice regime because, whether the result of a guilty plea or the judge’s decision after a trial, the sentence will be vulnerable only in the unlikely event the judge picks the wrong sentencing range. Thus, resources that today are devoted to sentence appeals—which can be numerous, especially in determinate sentencing systems¹³²—can be transferred to parole hearings and appeals of parole board decisions in a preventive justice regime. Appellate courts presumably would defer to board decisions in most cases, especially if the board relies on an RAI approved by the

128. 408 U.S. 471 (1972).

129. *Id.* at 489.

130. *Gagnon v. Scarpelli*, 411 U.S. 778, 788 (1973).

131. *See Flores v. Stanford*, 18 CV 2468 (VB), 2019 WL 4572703, at *11 (S.D.N.Y. Sept. 20, 2019) (holding that a due process claim is stated when “[p]arole [b]oard commissioners commonly do not even read such offenders’ files before conducting parole interviews or making parole determinations; make parole determinations based at least in part on a risk assessment algorithm the workings of which no defendant knows or understands; predetermine parole outcomes before conducting parole interviews; and pay no attention to other commissioners’ questions and offenders’ answers during interviews”).

132. WAYNE LAFAVE ET AL., 6 CRIMINAL PROCEDURE § 26.3(g) (4th ed. 2019) (noting that appeals of sentences in indeterminate regimes were rare, but that “[w]ith the advent of presumptive sentencing and guideline sentencing, the landscape of appellate sentencing review in this country was altered dramatically”).

relevant jurisdiction-wide authority. But they could reverse a board's decision when it is arbitrary and capricious because it clearly violates *Jackson* or the principles governing RAIs outlined in Parts III and IV below.

The final piece of the preventive justice puzzle is staffing the parole board and establishing its decision-making process. Since the sole function of the board is to assess risk as a predicate for determining whether release should occur and, if so, under what circumstances, the board should be composed primarily of experts on risk assessment and risk management. The type of board proposed by Rhine, Reitz, and Petersilia fits the bill. Their board would consist of people recommended by a nonpartisan panel and include expert criminologists and professionals versed in risk assessment.¹³³

To assist in that endeavor, the board must have access to results from RAIs. For the reasons already canvassed, this last feature is crucial.¹³⁴ Without RAIs, risk assessment is a guessing game. With RAIs, legal decision-makers can identify the small group of individuals who have a significant likelihood of committing violent crime in the future in the absence of confinement, with the aim of releasing everyone else at the end of their minimum sentence. But explaining why and how that would work requires a significant amount of background.

III. THE USEFULNESS OF RISK ASSESSMENT INSTRUMENTS

Despite their potential advantages, the risk algorithms used in the criminal justice system today are highly controversial. A common claim is that they are not good at what they purport to do, which is to identify who will offend and who will not, who will be responsive to rehabilitative efforts and who will not be. But the tools are also maligned as racially biased, dehumanizing, and, for good measure, antithetical to the foundational principles of criminal justice. A sampling of recent article and book titles illustrates the point: "Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control,"¹³⁵ "Risk as a Proxy for Race: The Dangers of Risk Assessment,"¹³⁶ and *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*.¹³⁷ In 2018, over 110 civil rights groups signed a statement calling for an end to

133. Rhine et al., *supra* note 113, at 283.

134. See discussion *supra* Subparts I & II.C.

135. Sarah Valentine, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control*, 46 *FORDHAM URB. L.J.* 364 (2019).

136. Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 *FED. SENT'G REP.* 237 (2015).

137. VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018).

pretrial RAIs.¹³⁸ That same year, twenty-seven academics stated that “technical problems” with RAIs “cannot be resolved.”¹³⁹ And in 2020, another group of 2,435 scholars from a wide range of disciplines “demanded” that the Springer publishing company “issue a statement condemning the use of criminal justice statistics to predict criminality” because of their unscientific nature.¹⁴⁰

These claims have some basis in fact, but they can easily be overblown. And if the impact of these criticisms is to prevent the criminal justice system from using algorithms, a potentially valuable means of reform will be lost. A key argument in favor of algorithms is comparative in nature. While algorithms can be associated with a number of problems, alternative predictive techniques may well be much worse in each of these respects. Unstructured decision-making by judges, parole officers, and mental health professionals is notoriously bad, biased, reflexive, and often reliant on stereotypes and generalizations that ignore the goals of the system.¹⁴¹ Algorithms can do better, at least if subject to certain constraints.

This Part of the Article addresses complaints that RAIs are unlikely to be useful at sentencing, while Part IV focuses on claims that, even if they are useful, they are unjust. After describing RAIs in more detail, this Part of the Article briefly describes the requisites they should meet to help in the postconviction setting. If risk is a legal issue, which it often is in today’s sentencing regimes and would definitely be in a preventive justice regime, RAIs could, in theory, provide crucial input for legal decision-makers. But, in fact, RAIs can only do so if the information they provide about risk is, to use evidentiary terminology, both “material” (logically related to the questions the law asks) and “probative” (sufficiently accurate about the information it provides). Further, and most importantly, if optimal value is to be obtained from such an RAI, its results should be treated as presumptively dispositive on the issue of risk.

138. THE LEADERSHIP CONF. EDUC. FUND, THE USE OF PRETRIAL “RISK ASSESSMENT” INSTRUMENTS: A SHARED STATEMENT OF CIVIL RIGHTS CONCERNS 1, 10 (2018), <http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf>.

138. CHELSEA BARABAS ET AL., TECHNICAL FLAWS OF PRETRIAL RISK ASSESSMENTS RAISE GRAVE CONCERNS 4–6 (2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf; Chelsea Barabas et al., *The Problems with Risk Assessment Tools*, N.Y. TIMES (July 17, 2019), <https://www.nytimes.com/2019/07/17/opinion/pretrial-ai.html>.

140. Coal. for Critical Tech., *Abolish the #TechtoPrisonPipeline*, MEDIUM (June 23, 2020), <https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16>.

141. Mirko Bagaric et al., *Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire*, 88 U. CIN. L. REV. 1037, 1057–58, 1063–67 (2020).

A. *The Nature of Risk Assessment Instruments*

John Monahan and Jennifer Skeem have provided a helpful typology of “risk assessment.”¹⁴² That term encompasses a number of different practices that differ in the extent to which they: (1) rely on empirically valid risk and protective factors, (2) use a structured method for measuring these risk and protective factors, (3) establish a procedure for combining scores on the individual risk and protective factors into a total score, and (4) produce a final estimate of risk. *Clinical* risk assessment—in many settings, the traditional and still-typical method used by many judges, parole boards, and mental health professionals—structures none of these components; rather, an estimate of risk is based on experience and perhaps intuition, and the factors considered may vary from case to case and be applied differently in different cases.¹⁴³ *Checklist* risk assessment provides structure on the first component by listing the factors that should be considered.¹⁴⁴ *Structured professional judgment* (“SPJ”) risk assessment satisfies the first two components by providing a list of factors and indicating how they should be measured (e.g., on a scale of 0–2) but avoids combining these measures for a total score, instead counselling that the item ratings be considered merely in arriving at an overall conclusion about risk.¹⁴⁵ *Adjusted actuarial* risk assessment lists the factors, describes how they should be measured, and produces a total score, but allows evaluators to adjust or modify the score based on clinical judgment that is not structured by the instrument.¹⁴⁶ Stand-alone *actuarial* risk assessment does not permit such adjustments, but rather produces a probability estimate that is considered final.¹⁴⁷

The term “risk assessment instruments” applies only to the last three types of practices.¹⁴⁸ Consider these four examples. The Violence Risk Appraisal Guide-Revised (“VRAG”), is an actuarial RAI relied on extensively in Canada and in several jurisdictions within the United States, mostly in connection with sentencing.¹⁴⁹ It

142. John Monahan & Jennifer Skeem, *Introductory Discussion of the Science—Research Approaches in Violence Risk Assessment*, in 2 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 9:11 (2019–2020 ed.).

143. *See id.*

144. *Id.*

145. *See id.*

146. *See* Christopher Slobogin, *Risk Assessment and Risk Management in Juvenile Justice*, 27 CRIM. JUST. 10, 12 (2013).

147. *See id.* at 11–12.

148. *See* Monahan & Skeem, *supra* note 142, § 9.11.

149. For a description of the instrument and relevant research, see Grant T. Harris et al., *Prospective Replication of the Violence Risk Appraisal Guide in Predicting Violent Recidivism Among Forensic Patients*, 26 LAW & HUM. BEHAV. 377, 378–79 (2002).

contains twelve largely static risk factors, having to do with the individual's score on the Psychopathy Checklist (a measure of psychopathy that takes into account criminal history), elementary school misconduct, diagnosis (with personality disorders positively correlated with risk and—perhaps surprisingly—schizophrenia negatively correlated), age, presence of parents in home before age sixteen, performance on conditional release, nonviolent offenses, marital status, victim injury, victim gender, and history of alcohol abuse.¹⁵⁰ The evaluator assigns a numerical subscore in connection with each risk factor and then adds the subscores to determine a total score that can range from less than negative twenty-one to more than twenty-eight.¹⁵¹ Initial research associated the lowest score on the VRAG with a 0 percent chance of violent offending within seven years, and the highest score with a 100 chance of violent offending within that period; seven other “bins” or ranges are associated with recidivism probabilities of 8 percent to 76 percent.¹⁵²

A much slimmer actuarial RAI, called the Non-Violent Risk Assessment (“NVRA”), is used in Virginia.¹⁵³ The NVRA actually consists of several RAIs, utilizing a separate tool for different categories of crime. For instance, if the crime of conviction is fraud, the instrument relies on just five static factors: offender age at the time of the offense (twenty-two points if the offender is younger than twenty-one), offender gender (ten points for males; one point for females), prior adult felony convictions (ten points if more than three convictions), prior adult incarcerations (four points if one to nine; thirty-two points if more than ten), and legal restraints (e.g., probation) at the time of conviction (six points).¹⁵⁴ If an offender receives more than thirty points on this instrument, prison is recommended; otherwise, the tool counsels the judge to consider “alternative punishment.” Note that on this instrument, a young male receives a prison recommendation if he has one prior felony conviction before the offense of conviction.¹⁵⁵

150. *Id.* at 378.

151. *See id.* at 385 tbl.2.

152. *See id.*

153. For a description of the original instruments and some data on their implementation, see Richard P. Kern & Meredith Farrar-Owens, *Sentencing Guidelines with Integrated Offender Risk Assessment*, 16 FED. SENT'G REP. 165, 169 (2004) (“The non-violent risk assessment tool adapted as part of the discretionary sentencing guidelines serves to safely divert a significant share of low-risk felons away from expensive prison beds into less costly alternative punishment programs.”).

154. BRANDON L. GARRETT ET AL., NONVIOLENT RISK ASSESSMENT IN VIRGINIA SENTENCING: THE SENTENCING COMMISSION DATA 15 (2018), <https://www.law.virginia.edu/system/files/news/spr18/UVA%20Law%20NVRA%20Sentencing%20Analysis%20and%20Judicial%20Survey%2C%20March%201%2C%202018.pdf>.

155. *Id.*

A third, more complicated instrument is the Correctional Offender Management Profiling for Alternative Sanctions tool (“COMPAS”).¹⁵⁶ It can be used either as an actuarial RAI or an adjusted actuarial RAI. Originally developed as an aid to corrections departments making decisions about placing, managing, and treating offenders, it has also been used to make sentencing, parole, and pretrial release decisions.¹⁵⁷ The COMPAS is much less transparent than the VRAG or the NVRA. Although the questions canvassed by the COMPAS (over 120) are available, the weight given to particular answers to those questions has not been disclosed, nor is it publicly known whether a given answer is even relevant to the ultimate risk assessment. The company that developed the COMPAS, Equivant (formerly Northpointe), has stated that the factors that affect the tool’s Violence Recidivism score consist of the offender’s age at the time of the assessment, the offender’s age at the time of first adjudication, the History of Violence Scale, the History of Noncompliance Scale, and the Vocational Educational Scale, but it is not willing to reveal the impact of each factor on the risk score or how the various scales are constructed.¹⁵⁸ Also, in contrast to the VRAG and the NVRA, the COMPAS recidivism scores are reported not in terms of probabilities or prison recommendations but in terms of “deciles;” inclusion in the first decile does not mean that the offender poses a 10 percent probability of recidivism, but rather that the offender fits in the bottom 10 percent of the group on which the COMPAS was validated.¹⁵⁹ Offenders are designated “low risk” if they fit within the first through fourth deciles, “medium risk” if within the fifth through seventh deciles, and “high risk” if within the eight through tenth deciles.¹⁶⁰

A final RAI that is sometimes used at sentencing, and often used to help figure out dispositional programs that might reduce risk, is the HCR-20, version 3 (“HCR-20”).¹⁶¹ As the name implies, this RAI consists of twenty risk factors, ten having to do with historical matters, five relating to clinical symptoms, and five relating to risk

156. For a PDF of one version of the instrument, see Northpointe, Inc., *Risk Assessment*, COMPAS CORE, www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html (last visited Mar. 15, 2021).

157. See JENNIFER L. SKEEM ET AL., ASSESSMENT OF EVIDENCE ON THE QUALITY OF THE CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS (COMPAS) 4 (2007).

158. See NORTHPOINTE, INC., PRACTITIONER’S GUIDE TO COMPAS 25 (2012), http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf.

159. *Id.* at 9.

160. *Id.* at 11.

161. For a description of this instrument, see Kevin S. Douglas & Christopher D. Webster, *The HCR-20 Violence Risk Assessment Scheme: Concurrent Validity in a Sample of Incarcerated Offenders*, 26 CRIM. JUST. & BEHAV. 3, 8 (1999).

management or treatment.¹⁶² The historical factors are previous violence, age at first violent incident, relationship instability, employment problems, substance use problems, major mental illness, psychopathy, early maladjustment, personality disorder, and prior supervision failure.¹⁶³ The clinical factors are all “dynamic” or changeable: lack of insight, negative attitudes, active symptoms of major mental illness, impulsivity, and unresponsiveness to treatment.¹⁶⁴ The risk management factors are also dynamic: unfeasibility of plans, exposure to destabilizers, lack of personal support, noncompliance with remediation attempts, and stress.¹⁶⁵ Each of the twenty factors is scored on a scale of zero to two, so that the maximum total score is forty.¹⁶⁶ However, the developers of the HCR-20 strongly counsel that a strictly mathematical assessment should be avoided and that, instead, individuals should simply be characterized by the evaluator as “high,” “medium,” or “low” risk.¹⁶⁷ Thus, the HCR-20 is generally considered an SPJ tool, as distinguished from instruments like the VRAG, NVRA, and COMPAS, which are actuarial.¹⁶⁸

All of these instruments except the HCR-20 rely on regression models, a relatively simple form of machine learning that involves humans inputting data and then relies on computers to calculate which variables correlate most strongly with the outcome variable of interest.¹⁶⁹ Developers of these instruments must make several decisions during the RAI construction process. They must decide where to get their data (e.g., government records, self-reports), what data to obtain with respect to reoffending (arrest, convictions, probation revocations, institutional incidents), what counts as a risk or protective factor, and the weight that the risk and protective factors will be assigned in the algorithm.¹⁷⁰

More sophisticated versions of RAIs involve computers setting their own rules about how much weight to assign variables, based on large amounts of “training” data.¹⁷¹ The inner workings of this type

162. *Id.*

163. *Id.*

164. *Id.*

165. *Id.*

166. *Id.*

167. See Kevin S. Douglas et al., *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography*, SIMON FRASER UNIV. 3, 27 (2005) https://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1362&context=psych_cmhsr.

168. See *id.*; Douglas & Webster, *supra* note 161, at 9.

169. See Kern & Farrar-Owens, *supra* note 153, at 176–78.

170. Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 67–86 (2017).

171. For a description of machine learning and risk assessment, see Richard Berk & Jordan Hyatt, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, 27 FED. SENT'G REP. 222 (2015).

of machine-learning RAI are much less transparent to human observers than the RAIs described above. While such RAIs are not currently widely used in the criminal justice system,¹⁷² they may become more available in the near future.

B. Ensuring the Materiality, or Fit, of RAIs

One of the reasons the developers of SPJ instruments like the HCR-20 avoid a more quantitative approach is the belief that associating probabilities with an individual offender—as might occur with tools like the VRAG and the NVRA—misrepresents the risk assessment enterprise to the extent it suggests that research about groups can predict whether a given individual will reoffend. The point was put most strongly by Stephen Hart, a developer of another SPJ instrument, who stated that “[i]t is impossible to directly measure (using some technology) or calculate (using some natural law) the specific probability or absolute likelihood that a particular offender will commit . . . violence, and even impossible to estimate this risk with any reasonable degree of scientific or professional certainty.”¹⁷³ David Cooke and Christine Michie have chimed in, stating—and purporting to demonstrate statistically—that “predictions of future offending cannot be achieved in the individual case with any degree of confidence.”¹⁷⁴ If that is true, the numerical conclusions reached by the VRAG and the other actuarial instruments described above are meaningless. Rather, all that can be said is that, given a person’s constellation of risk and protective factors, they are more or less likely to offend if certain interventions do not take place.

The view that nomothetic (group-based) information is irrelevant to a decision about an individual is not new with Hart, Cooke, and Michie. In 1942, Gordon Allport, a psychologist, said that:

A fatal nonsequitur occurs in the reasoning that if 80% of the delinquents who come from broken homes are recidivists, then this delinquent from a broken home has an 80% chance of becoming a recidivist. The truth of the matter is that this

172. TURGUT OZKAN, PREDICTING RECIDIVISM THROUGH MACHINE LEARNING 16 (2017) (unpublished Ph.D. dissertation, University of Texas at Dallas) (“[C]riminology as a field is lagging behind other branches of science when it comes to incorporating novel algorithmic decision-making tools.”).

173. Stephen Hart, *Evidence-Based Assessment of Risk for Sexual Violence*, 1 CHAPMAN J. CRIM. JUST. 143, 164 (2009).

174. See 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, 259 (2010).

delinquent has either 100% certainty of becoming a repeater or 100% certainty of going straight.¹⁷⁵

In 1994, Justice Coyne of the Minnesota Supreme Court wrote: “Not only are . . . statistics concerning the violent behavior of others irrelevant, but it seems to me wrong to confine any person on the basis not of that person’s own prior conduct but on the basis of statistical evidence regarding the behavior of other people.”¹⁷⁶ Consistent with these views, some courts have excluded “statistical speculation” about risk on the ground that it is not “individualized” or “particularized.”¹⁷⁷

David Faigman, John Monahan, and I have called the issue raised by these types of comments the “G2i” problem, the translation of general information to individual cases.¹⁷⁸ While we agree that the law has paid insufficient attention to the difficulties inherent in this translation process, we disagree with the notion that it is not possible. Cooke and Michie’s statistical argument that confidence intervals about recidivism probabilities are nonsensical has been roundly debunked by noted statisticians Peter Imrey and Philip Dawid, who found their position “seriously mistaken in many particulars” and concluded that their analysis should “play no role in reasoned discussions about violence recidivism risk assessment.”¹⁷⁹ As Karl Hanson and Philip Howard have noted, the declarations by Cooke, Michie, and Hart, “if true[,] . . . would be a serious challenge to the applicability of any empirically based risk procedure to any individual for anything.”¹⁸⁰

I would go further. If these assertions against nomothetic risk assessment are true, any type of expert testimony—empirically based or not—would be suspect. Even the expert who purports to bottom his or her conclusions solely on an interview with the individual in question relies—consciously or not—on stereotypes, past experiences with “similar” individuals, and lessons learned from literature about related groups. In the specific context of risk assessments, factors

175. GORDON W. ALLPORT, SOC. SCI. RSCH. COUNCIL, THE USE OF PERSONAL DOCUMENTS IN PSYCHOLOGICAL SCIENCE 156 (1942).

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

177. *See* *Porter v. Commonwealth*, 661 S.E.2d 415, 440 (Va. 2008) (excluding defense testimony on this ground); *see also* *United States v. Taylor*, 583 F. Supp. 2d 923, 942 (E.D. Tenn. 2008) (excluding defense testimony that “invites the jury to make decisions based upon group characteristics and assumptions”); *Rhodes v. State*, 896 N.E.2d 1193, 1194–96 (Ind. Ct. App. 2008) (holding that using an LSI-R score as an aggravating factor at sentencing was impermissible).

178. *See* David Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417, 417–18 (2014).

179. Peter B. Imrey & A. Philip Dawid, *A Commentary on Statistical Assessment of Violence Recidivism Risk*, 2 STAT. & PUB. POL’Y 1, 1 (2015).

180. R. Karl Hanson & Philip D. Howard, *Individual Confidence Intervals Do Not Inform Decision-Makers About the Accuracy of Risk Assessment Evaluations*, 34 LAW & HUM. BEHAV. 275, 277 (2010).

such as age, gender, criminal history, and personality traits influence (or do not influence) the evaluator's conclusions because they are "known" or intuited to have particular correlations with, or effects on, behavior. Barbara Underwood put it well in 1979:

Although the clinician need not identify in advance the characteristics he will regard as salient, he must nevertheless evaluate the applicant on the basis of a finite number of salient characteristics, and thus, like the statistical decisionmaker, he treats the application as a member of a class defined by those characteristics.¹⁸¹

This does not mean that a risk assessment based on an RAI should be blasé about the G2i issue. A risk estimate from an RAI (indeed, any conclusion about risk) should not purport to say that a particular offender has X probability of reoffending. Nor should it state that the person will, or will not, reoffend—a fact that, in almost all cases, *is* unknowable. Rather, an evaluator using an RAI to estimate the risk an offender poses should report that the offender received a risk score that is consistent with the scores of a group, X percent of which offended in the past. So phrased, a conclusion about risk offered in a legal proceeding is not "irrelevant" on G2i grounds.¹⁸²

However, such a conclusion could still be irrelevant on a number of other grounds. It is a constitutional axiom that the elements of crime be clearly delineated, in part to provide sufficient notice to the public about prohibited conduct, but primarily to control the discretion of police, prosecutors, and judges, who otherwise might abuse vague laws.¹⁸³ Because risk assessments can have the same impact on liberty as a conviction for crime, legislatures and courts should similarly be obligated to set forth, as a matter of law, the elements of risk that must be proven. As I have argued in other work,¹⁸⁴ those elements are: (1) the probability, preferably quantified

181. Barbara D. Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L.J. 1408, 1427 (1979).

182. See Erica Beecher-Monas, *Lost in Translation: Statistical Inference in Court*, 46 ARIZ. ST. L.J. 1057, 1092 (2014) (calling this type of formulation a "solution to the G2i problem in the courts [T]he instrument does not say . . . that the tested individual is 26% likely to recidivate. He is merely part of a group with an average recidivism rate of 26%. Some will recidivate more, some less. The inference about the tested individual's likelihood of recidivism is left to the factfinder.").

183. See generally John Calvin Jeffries, Jr., *Legality, Vagueness, and the Construction of Penal Statutes*, 71 VA. L. REV. 189 (1985) (discussing the issues inherent in vague laws).

184. See Christopher Slobogin, *Assessing the Risk of Offending Through Algorithms*, in THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 432, 434 (Woodrow Barfield ed., 2020); Christopher Slobogin, *Principles of Risk Assessment: Sentencing and Policing*, 15 OHIO ST. J. CRIM. L. 583, 587–89 (2018).

in terms of a group; (2) of a particular outcome; (3) within a particular time frame; (4) in the absence of a given intervention (usually incarceration). Every RAI should address these points for the legal setting in question.

C. *Ensuring the Validity of RAIs*

Assume that an RAI produces data that meets all of the fit requirements. It provides group probability estimates regarding outcomes, time frames, and interventions that fit the inquiries the law asks. Decision-makers still should not rely on these results unless they are reasonably accurate. What lawyers call accuracy or reliability, social scientists call validity.¹⁸⁵ There are numerous measures of validity. As applied to RAIs, they are all aimed at determining the extent to which an instrument does what it purports to do.¹⁸⁶

As a matter of legal precedent, the validity inquiry is most sensibly framed under Federal Rule of Evidence 702 and the analogous rules in virtually every state, which require experts to use “reliable principles and methods” that are “reliably applied.”¹⁸⁷ This language was meant to operationalize the U.S. Supreme Court’s decision in *Daubert v. Merrell Dow Pharmacies*,¹⁸⁸ which tells courts to ensure that the basis of expert testimony has been subjected to some type of verification process, such as empirical testing, the generation of error rates, and peer review through publication in accepted journals.¹⁸⁹ Many jurisdictions hold that Rule 702 and the *Daubert* criteria do not apply in the postconviction setting at issue here, because they were devised as rules of evidence meant to apply at trials.¹⁹⁰ But that stance is hard to fathom; if the *Daubert* test must be met before evidence can be presented in a tort suit (the context of the *Daubert* case), it should certainly apply where weeks, months, or years of jail or prison time are at stake.

If *Daubert*, or something like it, did apply to RAIs, it would require: (1) that the probabilities the instrument generates be stated with an acceptable level of confidence (calibration validity), (2) that the RAI do a passable job at distinguishing high risk from low risk

185. *Reliability and Validity*, WAC CLEARINGHOUSE, <https://wac.colostate.edu/resources/writing/guides/reliability-validity/> (last visited Mar. 15, 2021).

186. Slobogin, *Principles of Risk Assessment: Sentencing and Policing*, *supra* note 184, at 588–89.

187. FED. R. EVID. 702(c)–(d).

188. 509 U.S. 579 (1993).

189. *Id.* at 589–94.

190. *United States v. Slager*, 912 F.3d 224, 235 n.4 (4th Cir. 2019) (“Applying *Daubert* would run contrary to the sentencing court’s ‘wide discretion’ to determine the ‘sources and types of evidence,’ which are relevant for defendants’ individualized sentencing.”); *United States v. Fields*, 483 F.3d 313, 342 (5th Cir. 2007) (“No Circuit that we are aware of has applied *Daubert* to sentencing.”).

individuals (discriminant validity), (3) that the RAI be validated on a population similar to the population in question (external validity), (4) that it have satisfactory interrater consistency (implementation validity, or reliability), and (5) that information about all of these factors be periodically updated (current validity).¹⁹¹ Notice that, as expressed here, all of these criteria leave considerable wiggle room (“acceptable level of confidence,” “passable job at distinguishing risk,” “similar populations,” “satisfactory reliability,” and “periodically update”). No instrument can achieve perfect predictive validity (an umbrella term for all five types of validity), and the extent to which an instrument should approximate that goal may, again, vary with the legal setting.

The argument could be made that, even if an RAI is adequately valid in all of these respects, it is not *needed* to reach accurate results. Rule 702 holds that expert testimony is admissible only if it is based on “specialized knowledge” that is “helpful” to the trier of fact.¹⁹² If laypeople unaided by experts are just as good at prediction as RAIs, then Rule 702 counsels against permitting the results of RAIs to be considered. And the same outcome might be appropriate if experts using clinical skills unaided by RAIs could do as well or better.

One might think that the statistical expertise needed to put together an actuarial RAI or the familiarity with research literature that goes into the creation of SPJ instruments would be sufficient evidence of specialized knowledge. But if judges or lay-parole boards are just as able to figure out who will reoffend as evaluators using RAIs, then arguably that specialized knowledge is not helpful to the factfinder. One study purported to find just that. In 2018, Julia Dressel and Hany Farid conducted a study that led them to conclude that the COMPAS “is no more accurate or fair than predictions made by people with little or no criminal justice expertise”¹⁹³ Specifically, they found, based on a comparison of human predictions and COMPAS predictions for one thousand defendants, that humans were correct in about 62 percent of the cases and the COMPAS was correct in 65 percent of the cases.¹⁹⁴

Dressel and Farid’s study had one significant problem, however. Given the way it was conducted, the humans essentially functioned like algorithms—they were each shown fifty mini-vignettes that listed only a few features of the defendant, all of which have a robust statistical relationship with reoffending.¹⁹⁵ Further, after the

191. For a description of these measures of predictive validity, see Jay P. Singh, *Predictive Validity Performance Indicators in Violent Risk Assessment*, 31 BEHAV. SCI. & L. 8, 17–18 (2013).

192. FED. R. EVID. 702(a).

193. Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES 1, 3 (2018).

194. *Id.* at 2.

195. *Id.* at 1–2.

subjects registered their opinion about whether the person in the vignette would recidivate, they were immediately told whether they were right or wrong, feedback that a judge or parole board virtually never receives but which, in the study, “trained” the participants about the most pertinent traits and how they are related to recidivism.¹⁹⁶ As Sharad Goel and his colleagues concluded, “these boosted ‘human predictions’ are far removed from unaided human judgment”¹⁹⁷ A follow-up study by Zhiyan Lin and others found that when lay participants are *not* provided feedback, they do much more poorly than an RAI, even when they are given base rate information about the average rate of offending of the population in question.¹⁹⁸ The authors also found that when the information given the humans was “noisier” (that is, when a much richer set of facts was provided than the barebones list of traits Dressel and Farid used), the humans did barely better than chance, whereas the statistical model the authors created was much better at distinguishing recidivists and non-recidivists.¹⁹⁹

These types of results replicate a large number of other studies finding that algorithms typically outperform human predictions, whether made by laypeople or trained clinicians.²⁰⁰ For instance, a 2006 meta-analysis of forty-one studies found that actuarial techniques routinely did better than clinical methods in every area investigated, and that with respect to predicting violent or criminal

196. Zhiyuan “Jerry” Lin et al., *The Limits of Human Predictions of Recidivism*, 6 SCI. ADVANCES 1, 1 (2020).

197. Sharad Goel, Ravi Shroff, Jennifer Skeem & Christopher Slobogin, *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment* 5 (Dec. 26, 2018) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3306723#.

198. Lin et al., *supra* note 196, at 2–4.

199. *Id.* at 5 (“[W]e also found that algorithms tended to outperform humans in settings where decision-makers have access to extensive information and do not receive immediate feedback and base rates are far from balanced, features of many real-world scenarios.”).

200. See Sarah J. Desmarais et al., *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.”); William M. Grove et al., *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 12 PSYCH. ASSESSMENT 19, 19 (2000) (finding that actuarial predictions are about 10 percent more accurate than clinical predictions); R. Karl Hanson & Kelly E. Morton-Bourgon, *The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies*, 21 PSYCH. ASSESSMENT 1, 6 (2009) (“For the prediction of sexual or violent recidivism, the actuarial measures designed for violent recidivism . . . were superior to any of the other methods.”); see also Green & Chen, *supra* note 38, at 90 (presenting the results of an experimental study in which human subjects “underperformed the risk assessment even when presented with its predictions . . .”).

behavior in particular, the actuarial approach was “clearly superior to the clinical approach.”²⁰¹ The study also found that even subsets of “best professionals” designated as experts did not outperform statistical formulae.²⁰² Several studies that compare algorithms to judges, clinicians, and correctional officers obtain similar results, probably because, despite their official position, the decisions of these people are like those of other people—largely intuitive, heuristic-based, subject to bias, and inattentive to base rates.²⁰³

201. Stefania Ægisdóttir et al., *The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction*, 34 COUNSELING PSYCH. 341, 368 (2006) (“Out of 1,000 predictions of violence, the statistical method should correctly identify [ninety] more violent clients than will the clinical method.”).

202. *Id.* (“We found that when clinicians were given the same or more information than the statistical formula, the formula did better.”).

203. See Thomas H. Cohen et al., *Examining Overrides of Risk Classifications for Offenders on Federal Supervision*, 80 FED. PROB. 12, 20–21 (2016) (finding that overrides by probation officers were “almost all . . . an upward adjustment” and that these overrides “demonstrated a weaker correlation between the adjusted risk levels and recidivism compared to the original risk levels”); Hanson & Morton-Bourgon, *supra* note 200, at 7 (finding that risk ratings that incorporated the professional judgments of probation officers, psychologists, and correctional staff “showed lower predictive accuracy than did the unadjusted actuarial scores”); Kleinberg et al., *supra* note 38, at 240–41 (finding that the use of algorithms in pretrial detention determinations would reduce crime rates by assisting human decision-makers); J. Stephen Wormith et al., *The Predictive Validity of a General Risk/Needs Assessment Inventory on Sexual Offender Recidivism and an Exploration of the Professional Override*, 39 CRIM. JUST. & BEHAV. 1511, 1534 (2012) (finding that a professional override by probation officers, psychologists, or social workers on the result suggested by an RAI designed to test sexual offense recidivism “led to a slight, but systematic, deterioration in the predictive validity . . .”); Jongbin Jung et al., *Simple Rules for Complex Decisions* 4 (Stan. Univ., Working Paper No. 1702.04690, 2017), <https://arxiv.org/pdf/1702.04690.pdf> (finding that when compared to judges in a pretrial setting a simple algorithm that only looked at two factors of arrestees, age and previous failures to appear, “consistently outperform[ed] the human decision-makers”); Kathleen Spencer Gore, *Adjusted Actuarial Assessment of Sex Offenders: The Impact of Clinical Overrides on Predictive Accuracy Gain* 63–64 (2007) (Ph.D. dissertation, Iowa State University) (on file with Iowa State University Digital Repository), <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=16536&context=rtd> (finding that psychologist overrides on the results produced by an instrument designed to predict recidivism for sex offenders “failed to even nominally exceed the [instrument] in terms of overall predictive accuracy”); see also Stephen D. Gottfredson & Laura J. Moriarty, *Clinical Versus Actuarial Judgments in Criminal Justice Decisions: Should One Replace the Other?*, 70 FED. PROB. 15, 15, 17 (2006) (stating that over-reliance on human judgment may undermine the accuracy of risk assessment because probation and parole officers may “concentrate on information that is demonstrably not predictive of offender behavioral outcomes . . .”).

D. *The Need for a Presumption in Favor of RAI Results*

Given their superiority to lay judgment, the results of an RAI that meets the fit and validity requirements discussed above should be given presumptive effect. Unfortunately, that rarely occurs in those jurisdictions that use RAIs. As one judge remarked about RAIs, “[i]t’s important to understand that it’s just a tool and that judges are the definitive answer.”²⁰⁴ The judicial decisions that have analyzed the use of RAIs at sentencing have likewise emphasized that the results of an RAI are but one factor to consider and should not be dispositive.²⁰⁵

Judges and parole boards are clearly the ultimate decision-makers about offender risk. But they should be aware that evaluator, judicial, and parole board adjustments to an RAI usually do not improve on the actuarial assessment. In fact, consistent with the studies comparing actuarial and clinical judgment, several studies find that professional “overrides” of an RAI’s risk estimate, whether by judges, probation officers, or other correctional professionals, *decrease* accuracy in predicting offending.²⁰⁶ For example, based on a sample of 3,646 offenders, Jean-Pierre Guay and Genevieve Parent found that adjustments to an RAI result made by probation officers were significantly less accurate than the unadjusted RAI.²⁰⁷ A study by Schmidt and others found that professional overrides decreased predictive validity and usually increased risk level.²⁰⁸ The most recent study likewise found that overrides typically result in an “upward reshuffling” of risk and a loss of predictive accuracy.²⁰⁹ The pretrial setting is no different. One study found that judges agreed with 84 percent of the RAI’s “detain” recommendations but only 47 percent of its “release” recommendations.²¹⁰

204. Matthew DeMichele et al., *The Intuitive Override Model: Nudging Judges Toward Pretrial Risk Assessment Instruments* 18 (2018) (draft report), <https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/5-Intuitive-Override-Model.pdf>.

205. See, e.g., *Malenchik v. State*, 928 N.E.2d 564, 573 (Ind. 2010); *State v. Loomis*, 881 N.W.2d 749, 769 (Wis. 2016).

206. See sources cited *supra* note 203.

207. Jean-Pierre Guay & Genevieve 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, 97 (2018).

208. Fred Schmidt et al., *Predictive Validity of the Youth Level of Service/Case Management Inventory with Youth Who Have Committed Sexual and Non-Sexual Offenses: The Utility of Professional Override*, 43 CRIM. JUST. & BEHAV. 413, 413 (2016).

209. Thomas H. Cohen et al., *Risk Assessment Overrides: Shuffling the Risk Deck Without Any Improvements in Prediction*, 47 CRIM. JUST. & BEHAV. 1609, 1624 (2020).

210. HUM. RTS. WATCH, “NOT IN IT FOR JUSTICE:” HOW CALIFORNIA’S PRETRIAL DETENTION AND BAIL SYSTEM UNFAIRLY PUNISHES POOR PEOPLE 92 (2017),

There are likely several explanations for these types of findings. Adjustments to RAI results may be based on unverified speculation about the traits that might affect risk, a belief that “special circumstances” (e.g., contriteness or surliness) warrant ignoring the risk score, or simple mistrust of quantified decision-making.²¹¹ Or they may stem from extraneous considerations. In particular, decision-makers know that a false negative decision, which results in release, is much more likely to be discovered than a false positive decision that results in incarceration; moreover, of course, these types of errors are much more likely to have professional and societal consequences.²¹² Evaluators, judges, and parole board members might also dislike the idea of surrendering a significant amount of their discretion to a table; as one Virginia judge put it, “I don’t do voodoo.”²¹³

Nonetheless, if the Rule 702/*Daubert* reliability standard is taken seriously, case-by-case modifications should be rare, based on unique circumstances or obvious anomalies (such as when a person designated as high risk for violence has since become disabled,²¹⁴ or when a person considered low risk voices a genuine threat of harm).²¹⁵ Some courts appear to recognize that fact.²¹⁶ It is important to

<https://www.hrw.org/report/2017/04/11/not-it-justice/how-californias-pretrial-detention-and-bail-system-unfairly>.

211. R. Karl Hanson, *What Do We Know About Sex Offender Risk Assessment*, 4 PSYCH., PUB. POL’Y & L. 50, 65 (1998) (recommending that clinicians be “exceedingly cautious” in making adjustments, but noting that “[t]hose skeptical of actuarial predictions will always find reasons to adjust actuarial estimates”).

212. Katherine E. McCallum et al., *The Influence of Risk Assessment Instrument Scores on Evaluators’ Risk Opinions and Sexual Offender Containment Recommendations*, 44 CRIM. JUST. & BEHAV. 1213, 1214 (2017).

213. Anne Metz et al., *Valid or Voodoo: A Qualitative Study of Attorney Attitudes Towards Risk Assessment in Sentencing and Plea Bargaining* 16 (Virginia Public Law and Legal Theory Research Paper No. 2020-25, 2020), <https://ssrn.com/abstract=3552018>.

214. See Daniel A. Krauss & Nicholas Scurich, *Risk Assessment in the Law: Legal Admissibility, Scientific Validity, and Some Disparities between Research and Practice*, 31 BEHAV. SCI. & L. 215, 221 (2013) (noting that an actuarial algorithm cannot incorporate unique risk factors, such as “if the offender incurred a physically incapacitating injury”).

215. Cf. *Kansas v. Hendricks*, 521 U.S. 346, 355 (1997) (noting that Hendricks stated that “the only sure way he could keep from sexually abusing children in the future was ‘to die’”).

216. *Malenchik v. State*, 928 N.E.2d 564, 573 (Ind. 2010) (“Having been determined to be statistically valid, reliable, and effective in forecasting recidivism, the assessment tool scores may, and if possible should, be considered to supplement and enhance a judge’s evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant.”); *State v. Loomis*, 881 N.W.2d 749, 758 (Wis. 2016) (contrasting “evidence-based sentencing” with “ad hoc decision

remember that carefully constructed RAIs are the products of considering and discarding scores of variables during development. The authors of the VRAG, for instance, analyzed approximately fifty possible risk and protective factors before paring them down to twelve.²¹⁷ Some of these variables were rejected for practical reasons (e.g., the difficulty of finding the relevant data) rather than empirical ones (e.g., no correlation with risk).²¹⁸ Adjustments based on the former variables should be fair game for litigants. But relying on a variable that was found to have no predictive value to change a person's risk score does not make empirical sense.

Of particular concern is the common finding that upward adjustments are more likely than downward adjustments and that these adjustments increase inaccuracy.²¹⁹ One of the common mistakes in this regard is to “double-count” criminal history. For instance, an RAI might indicate that an offender belongs in a low-risk category, but a judge, noting that the offender has committed two prior offenses, may decide otherwise; research in Virginia confirms that this is a common occurrence.²²⁰ Yet, since every RAI already incorporates criminal history, this assessment will almost certainly be erroneous.

IV. A DEFENSE OF RISK ASSESSMENT INSTRUMENTS AGAINST INJUSTICE CLAIMS

If they have sufficient fit and meet basic indicia of validity, RAIs, applied presumptively, can be very helpful to the legal system, especially if decarceration is a goal. But even if those conditions are met, RAIs have at least three more hurdles to clear. While critics of risk algorithms in the criminal justice system have often complained about “off-label” uses of RAIs (the fit problem) or the difficulty of getting risk right (the validity problem), they have aimed their

making” and stating that “Wisconsin’s commitment to evidence-based practices has been well documented”).

217. Marnie E. Rice & Grant T. Harris, *Violence Risk Appraisal Guide (VRAG)*, in *ENCYCLOPEDIA OF PSYCHOLOGY AND LAW* 847, 847 (Brian L. Cutler ed., 2008).

218. See Angèle Christin et al., *Courts and Predictive Algorithms* 4 (Oct. 27, 2015) (unpublished manuscript), http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf (noting that “dealing with a small sample size or a large amount of data” could make the model less accurate).

219. Wormith et al., *supra* note 203, at 1525.

220. Anne Metz et al., *Risk and Resources: A Qualitative Perspective on Low-Level Sentencing in Virginia*, *J. CMTY. PSYCHOL.* 1476, 1483 (2019) (finding that 83 percent of judges “consider[ed] the results of the NVRA as a ‘validating data point’ but not . . . ‘dispositive,’” because they also consider “the facts of the case, and the defendant’s criminal history”); see also Green & Chen, *supra* note 38, at 3 (finding that “participants responding to particularly salient features that are unevenly distributed by race (such as number of prior convictions)—essentially double-count[ed] features for which the risk assessment had already accounted”).

heaviest artillery at what they perceive to be the unfairness of using actuarial tools to determine risk.²²¹ Even if RAIs were to satisfy the most stringent fit and validity requirements, critics argue, their use should be avoided or minimized because they exacerbate racial and other biases, justify liberty deprivations using factors that are irrelevant to blameworthiness, and allow incarceration to be based on dehumanizing, nonindividualized determinations.²²² In the following discussion, these claims are labeled the egalitarian injustice, retributive injustice, and procedural injustice critiques.

These critiques have important implications for the jurisprudence of risk. But none of them should sound the death knell for RAIs. Each can be countered through careful construction and use of risk assessment tools and transparency about their inner workings.

A. *The Claim of Egalitarian Injustice*

As a matter of formal legal doctrine, the Equal Protection Clause of the Fourteenth Amendment is the focal point of discussion about the extent to which the government must treat its citizens equitably. Generally, the Clause has been construed to permit government statutes and practices that discriminate between similarly situated individuals as long as they are based on some rational reason, which need not be empirically supported but rather can rest simply on “common sense.”²²³ However, certain legal classifications—most prominently race and sex, but also including alienage and religion—are “suspect” because of their historical association with political and social oppression. Thus, a statute or practice that discriminates on those grounds is subject to “strict scrutiny” (in the case of race) or “intermediate scrutiny” (in the case of sex).²²⁴ The state may discriminate on the basis of race only if it can demonstrate that it has a compelling reason for doing so, using a means narrowly tailored to meet its objective, and discrimination on the basis of gender is permitted only upon a similar showing, albeit not quite as demanding.²²⁵

This interpretation of the Equal Protection Clause is usually called “anticlassification,” because it severely limits the government’s ability to discriminate or classify on the basis of a person’s race and sex, regardless of which race or gender is at issue. It is often distinguished from the “antisubordination” approach to equal protection, which focuses on whether the government is

221. See *infra* notes 244–45 and accompanying text.

222. See Green & Chen, *supra* note 38, at 1, 11, 16.

223. *FCC v. Beach Commc’ns, Inc.*, 508 U.S. 307, 313 (1993) (noting under rational basis review, a law “must be upheld against equal protection challenge if there is any reasonably conceivable state of facts that could provide a rational basis for the classification” made by the law).

224. See *Clark v. Jeter*, 486 U.S. 456, 461 (1988).

225. *Id.*

discriminating against a group that has historically been oppressed or subordinated to more privileged groups; under this approach, and in contrast to the anticlassification principle, race and sex may be taken into account to the extent that doing so will rectify wrongs against people of color and women.²²⁶ *Brown v. Board of Education*,²²⁷ which famously found “separate but equal” white and black schools to be a violation of equal protection, could be seen as an illustration of either the anticlassification or the antisubordination principle; under the educational system declared unconstitutional in the case, schools were classified on the basis of race, and that classification significantly disadvantaged a race that had been enslaved and then subject to Jim Crow laws.²²⁸ When the Court has been confronted with cases that required a choice between these two equal protection theories, however, it has usually chosen the anticlassification approach.²²⁹ For instance, affirmative action programs for black students could be seen as an implementation of the antisubordination version of equal protection, but the Court over time has held that, under most circumstances, they are unconstitutional because they classify on the basis of race. The Court’s current position is summed up in Chief Justice Roberts’ statement that “[t]he way to stop discrimination on the basis of race is to stop discriminating on the basis of race.”²³⁰

When a statute or practice does not explicitly differentiate on the basis of race or sex, but rather simply has a disparate racial or sex-based impact, it is usually not considered violative of equal protection

226. On the distinctions between anticlassification and antisubordination approaches see Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIA. L. REV. 9, 9–10 (2003).

227. 347 U.S. 483, 495 (1954).

228. See *id.*; Reva B. Siegel, *Equality Talk: Antisubordination and Anticlassification Values in Constitutional Struggles over Brown*, 117 HARV. L. REV. 1470, 1484–99 (2004) (noting that, immediately after *Brown*, some scholars interpreted it consistently with the anticlassification theory and others with antisubordination theory).

229. Kimberly Jenkins Robinson, *The Constitutional Future of Race-Neutral Efforts to Achieve Diversity and Avoid Racial Isolation in Elementary and Secondary Schools*, 50 B.C. L. REV. 277, 315 (2009) (“The Court’s current approach to equal protection, which has been labeled an antidiscrimination, anticlassification, or color-blind approach, emphasizes the impropriety of government use of racial classifications.”).

230. *Parents Involved in Cmty. Schs. v. Seattle Sch. Dist. No. 1*, 551 U.S. 701, 748 (2007) (plurality opinion); see also *Grutter v. Bollinger*, 539 U.S. 306, 329–30 (2003) (stating that enrolling a specified percentage of a particular group merely because of its race or ethnic origin “would amount to outright racial balancing, which is patently unconstitutional”). But see *Fisher v. Univ. of Tex. at Austin*, 570 U.S. 297, 314–15 (2013) (holding that considering race in a holistic methodology is permissible but must withstand strict scrutiny).

unless animus toward the racial or sex-based group is shown.²³¹ As the Supreme Court put it, “our cases have not embraced the proposition that a law or other official act, without regard to whether it reflects a racially discriminatory purpose, is unconstitutional solely because it has a racially disproportionate impact.”²³² This test recognizes that the absence of an explicit race or sex classification usually passes constitutional muster.

Application of these concepts to RAIs is not an easy task, but the overall conclusion should be that, with a few caveats, such instruments are not violative of equal protection if they provide relevant and probative results. A number of courts, including the Supreme Court, have held that protecting the public from dangerous individuals is a compelling state interest.²³³ On that assumption, traditional anticlassification theory might permit even explicit use of race in a risk algorithm if it has significant predictive validity and if alternatives to using it could not achieve the state’s aims as effectively. However, race alone is not a particularly strong predictor.²³⁴ In any event, in the 2017 decision of *Buck v. Davis*,²³⁵ a capital sentencing case, the Supreme Court declared, without even mentioning any possible state interest that might thereby be undermined, that “[i]t would be patently unconstitutional for a state

231. *Washington v. Davis*, 426 U.S. 229, 239 (1976).

232. *Id.*

233. *See United States v. Salerno*, 481 U.S. 739, 744 (1987) (stating that in the pretrial detention context, the federal government has “compelling interests in public safety . . .”); *Schall v. Martin*, 467 U.S. 253, 264 (1984) (stating the same and collecting cases standing for the proposition that “[t]he ‘legitimate and compelling state interest’ in protecting the community from crime cannot be doubted” and is “a weighty social objective . . .”) (internal citations omitted). Lower courts have found the prevention of even relatively minor crimes or speculative risks to be a compelling interest. *United States v. Chappell*, 691 F.3d 388, 399 (4th Cir. 2012) (“The Virginia statute serves the important—indeed compelling—interest of promoting public safety by prohibiting an individual from intentionally impersonating a peace officer for a broad range of improper aims . . .”); *Kachalsky v. Cacace*, 817 F. Supp. 2d 235, 271 (S.D.N.Y. 2011) (“[T]he state has an important government interest in promoting public safety and preventing crime.”); *May v. Hunter*, 451 F. Supp. 2d 1084, 1088 (C.D. Cal. 2006) (stating that “the state has a compelling interest to protect the public from those prisoners who are not in remission, and that in order to protect the public, it is necessary to provide those prisoners continuing mental health treatment until the underlying condition can be kept in remission”) (internal quotation marks omitted).

234. Michael T. Kirkpatrick & Margaret B. Kwoka, *Title VI Disparate Impact Claims Would Not Hurt National Security—A Response to Paul Taylor*, 46 HARV. J. ON LEGIS. 503, 524–25 (2009) (“Although there are data showing correlations between arrest rates and race, and incarceration rates and race,” no data demonstrate “either a general or a circumstantial correlation between race and crime.”) (internal quotation mark omitted).

235. 137 S. Ct. 759 (2017).

to argue that a defendant is liable to be a future danger because of his race.”²³⁶ Perhaps anticipating that conclusion, no mainstream RAI incorporates race into its algorithm.²³⁷ Nor, with one notable exception, do RAIs use ethnicity, alienage, or religion, which are also suspect classes and closely aligned with race in many circumstances.²³⁸

Several RAIs do, however, use gender to help differentiate risk levels.²³⁹ Not clear, post-*Davis*, is how the Court would react to that fact. Absent a Supreme Court case that directly addresses the issue, under current doctrine sex can only be used as a risk factor if there is a strong justification for doing so (as one lower court decision, discussed below, has found).²⁴⁰

If, instead, an RAI uses neither race nor sex as a risk factor but still produces results that have a disparate racial or gender impact, then formal classification is not occurring; in that case, use of the RAI is permissible if there is any rational basis for doing so, unless a discriminatory purpose can be shown. While the Supreme Court has not always required serious animus in its disparate impact jurisprudence, it has tended to require strong proof of discriminatory purpose in criminal cases.²⁴¹ In any event, developers of RAIs are not likely to have harbored or intended to implement animus toward any given racial group, and in fact probably want to avoid disproportionate outcomes.²⁴² Thus, a disparate impact argument against RAIs is unlikely to prevail.

236. *Id.* at 775.

237. CTR. ON RACE, INEQ. & THE LAW, N.Y. UNIV. SCH. OF L. & AM. C.L. UNION, WHAT DOES FAIRNESS LOOK LIKE? CONVERSATIONS ON RACE, RISK ASSESSMENT TOOLS, AND PRETRIAL JUSTICE 12–13 (2018), <http://www.law.nyu.edu/sites/default/files/Final%20Report--ACLU-NYU%20CRIL%20Convening%20on%20Race%20Risk%20Assessment%20%20Fairness.pdf> (addressing the absence of race in RAI algorithms).

238. Thomas H. Cohen et al., *Revalidating the Federal Pretrial Risk Assessment (PTRA): A Research Summary*, 82 FED. PROB. 23, 23–24 (2018) (noting that the PTRA includes “citizenship” as a variable).

239. *See, e.g.*, Richard P. Kern & Meredith Farrar-Owens, *Sentencing Guidelines with Integrated Offender Risk Assessment*, 16 FED. SENT’G REP. 165, 166 (2004) (discussing the use of gender as an RAI factor).

240. *See infra* Subpart IV.A.2.

241. Michael Coenen, *Spillover Across Remedies*, 98 MINN. L. REV. 1211, 1229–30 (2014) (“[T]he discriminatory purpose rule applies with equal force in *criminal* cases” and noting, for instance, that “notwithstanding powerful” evidence of discrimination against black defendants in cocaine prosecutions, “*Washington v. Davis* and its progeny stood as an impenetrable barrier to equal protection relief, as black crack-cocaine defendants never managed to gather enough evidence to satisfy the discriminatory purpose requirement.”).

242. Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1193 (2017).

A number of writers have nonetheless insisted that RAIs violate related subconstitutional norms when they have a disparate racial impact,²⁴³ and some have argued that these norms are also violated when there is disparate impact on other grounds, such as poverty.²⁴⁴ Surfacing these concerns and responding to them is an important aspect of exploring the justness of algorithmic risk assessment. The overall conclusion reached here is that, given the importance of valid RAIs to achieving the state's goals of protecting the public and reducing incarceration, neither formal equal protection doctrine nor other inequality concerns should derail development of RAIs.

However, the balancing of state and individual interests of the type that is inherent in equal protection analysis does trigger fairness concerns. Specifically, it raises questions about whether certain outcome variables—most predominately, certain types of criminal history—are out-of-bounds on race discrimination grounds, whether protected characteristics—particularly sex—can nonetheless be explicitly considered in developing RAIs, and whether other characteristics—such as being unemployed—are so weakly predictive that they cannot survive even the minimal rational basis test.

1. Race

Probably the best-known critique of RAIs came in a 2016 article published in *ProPublica* about the COMPAS.²⁴⁵ The article reported that the instrument was “biased against blacks” because it produced a false positive rate for that group that was twice the false positive rate for white people when the cut-off was “medium” risk (45 percent to 23 percent) and three times the white false positive rate for those rated “high” risk (16 percent to 5 percent).²⁴⁶ The developer of the COMPAS did not disagree with these conclusions but pointed out that among both black and white people, groups with the same scores recidivated at the same rates, with, for instance, 21 percent of the high risk group for black people and 17 percent of the high risk group for white people reoffending.²⁴⁷ Further, as other commentators

243. Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1025 (2017) (arguing for “algorithmic affirmative action”); Harcourt, *supra* note 136, at 237.

244. See Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 805–06 (2014).

245. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

246. Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

247. Sam Corbett-Davies et al., *What Makes an Algorithm Fair?*, MEDIUM (Oct. 28, 2016), <https://medium.com/soal-food/what-makes-an-algorithm-fair-6ad64d75dd0c> (pointing out that “[d]efendants assigned the highest risk score

pointed out, the differential false positive rates produced by a reasonably well-calibrated instrument like the COMPAS are inevitable as a statistical matter when, as the data from the sample used by *ProPublica* seemed to indicate,²⁴⁸ the black crime rate is higher than the white crime rate. When black defendants have a higher rate of reoffending, a well-calibrated algorithm will classify a greater proportion of black defendants as high risk, and thus a greater proportion of black defendants who ultimately do not reoffend will have been classified as high risk.²⁴⁹

The COMPAS abides by the anticlassification principle; race is not considered. Nonetheless, it has been lambasted by a number of commentators. For instance, the RAND Corporation stated:

By accurately reflecting base rate criminality and thereby treating black people as higher risk than white people, COMPAS perpetuates a disproportionate impact along racial lines. . . . If existing base rates are unjust (because of historical factors and structural racism), it can be argued that accurate algorithms, such as COMPAS, are complicit in and contribute to this society-wide injustice.²⁵⁰

This language mischaracterizes the interaction of race and risk on the COMPAS in several ways. First, the algorithm does not “treat[] black people as higher risk than white people.” It treats people with risk factors X, Y, and Z as higher risk than those who do not have those risk factors—a greater proportion of whom happen to be black. Second, its impact is not “disproportionate” to the amount

[on the COMPAS] reoffended at almost four times the rate as those assigned the lowest score (81 percent versus 22 percent),” and that “among defendants who scored a seven on the COMPAS scale, 60 percent of white defendants reoffended, which is nearly identical to the 61 percent of black defendants who reoffended”); Larson et al., *supra* note 246 (“Black defendants who were classified as a higher risk of violent recidivism did recidivate at a slightly higher rate than white defendants (21 percent vs. 17 percent) . . .”).

248. Larson et al., *supra* note 246 (“Across every risk category, black defendants recidivated at higher rates.”).

249. Corbett-Davies et al., *supra* note 247 (“If the recidivism rate for white and black defendants is the same within each risk category, and if black defendants have a higher overall recidivism rate, then a greater share of black defendants will be classified as high risk. And if a greater share of black defendants are classified as high risk, then . . . a greater share of black defendants who do not reoffend will also be classified as high risk.”). Note that the same phenomenon would occur, for instance, with age and sex; because young people and males commit more crimes, a well-calibrated RAI will produce more false positives at each score among youth and men than among older people and women.

250. OSONDE A. OSOBA ET AL., ALGORITHMIC EQUITY: A FRAMEWORK FOR SOCIAL APPLICATIONS 20 (2019); *see also* MICHAEL TONRY, DOING JUSTICE, PREVENTING CRIME 180–82 (2020).

of crime black individuals commit; as RAND admits, the COMPAS accurately reflects base rate criminality. Third, if that base rate criminality is unjust because of structural racism, risk assessment is not the central problem; rather the entire criminal justice system, which assumes people can be punished for their crimes, is “complicit in and contributes to this society-wide injustice.” While there is a strong basis for calling the black crime base rate “unjust” from a sociological point of view, that critique, carried to its logical limit, would end far more than risk assessment; indeed, it is the predicate for the Abolish Prison movement.²⁵¹

ProPublica’s less radical solution to the problem—equalizing false positives rates for both races—has its own difficulties. This approach, sometimes called *classification parity*,²⁵² would mean that more white people would be classified as high risk when in fact they are not, or that more people of color who are high risk would be classified as low risk, or both. Additionally, as Richard Berk has suggested, that route would probably result in more victims of color, as individuals who should have been detained are released back into their communities.²⁵³

The approach taken by the COMPAS, which tries to ensure that people who pose the same degree of risk are in the same risk category, regardless of race, is a preferable method of ensuring fairness.²⁵⁴ First, it sidesteps possible concerns about violating the Equal Protection Clause by avoiding intentional misclassification of white or black people. Second, it avoids the political blowback that is likely to accompany migrating affirmative action principles to the criminal context. Third, and most importantly, it treats people of equal risk

251. See PAUL BUTLER, *CHOKEHOLD* 229–30 (2018) (“U.S. prisons are built for black men, and black men will be free, literally and figuratively, only when prisons are no more. . . . Think of abolition as the third gift people who fight for African American freedom will have provided the country, after they defeated slavery and Jim Crow.”).

252. SAM CORBETT-DAVIES & SHARAD GOEL, *THE MEASURE AND MISMEASURE OF FAIRNESS: A CRITICAL REVIEW OF FAIR MACHINE LEARNING* 11 (2018), <https://arxiv.org/pdf/1808.00023.pdf>.

253. Richard Berk, *Accuracy and Fairness for Juvenile Justice Risk Assessments*, 16 J. EMPIRICAL LEGAL STUD. 175, 191 (2019) (“When one adjusts algorithms to make them more fair for black perpetrators, one risks increasing unfairness for black crime victims.”).

254. See Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2272 (2019) (“[A]lgorithmic affirmative action, in essence, constitutes a rejection of actuarial risk assessment itself.”); Gina Vincent & Jodi Viljoen, *Racist Algorithms or Systemic Problems? Risk Assessments and Racial Disparities*, 47 CRIM. JUST. & BEHAV. 1576, 1580 (2020) (describing the views of several empiricists on the issue of racial bias and concluding “by our ethical standards, there is currently no valid evidence that instruments in general are biased against individuals of color. Where bias has been found, it appears to have more to do with the specific risk instrument.”).

equally. Each risk classification represents the same risk, regardless of race. *Propublica's* antisubordinationist approach clearly violates not just anticlassification principles but also calibration (accuracy) mandates.

At the same time, caution is necessary in evaluating the racial impact of RAIs. For instance, it is well-documented that people of color are more likely to be arrested for drug-possession crime than white people who engage in the same behavior, in part because of biased policing practices and in part because such crimes, when committed by poorer individuals, are more likely to be publicly observable.²⁵⁵ The same is probably also true of most misdemeanors.²⁵⁶ If an RAI uses *arrests* for these types of crimes as an outcome variable, as the COMPAS appears to do, it will be predicting that people of color commit more of these crimes than white people, when in fact they do not.²⁵⁷ Sandra Mayson makes the point with an anecdote: in New Orleans, where she practiced law, a black defendant with three arrests was not much of a concern—but a white defendant with three arrests was “really bad news!”²⁵⁸

There are two solutions to this problem. The first is to change the outcome variable to either convictions rather than arrests (on the assumption that convictions more accurately indicate criminal activity between races), or to arrests (or convictions) for violent crimes only. Changing the variable to violent crimes is preferable, not only for empirical reasons (arrests and convictions for violent crimes are less likely to be racially contingent)²⁵⁹ but because refusing to release

255. See Chelsea Barabas et al., *Technical Flaws of Pretrial Risk Assessment Tools Raise Grave Concerns*, MIT MEDIA LAB 3 (July 17, 2019), https://damprod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf; Peter Reuter et al., *Assessing the Crack-Down on Marijuana in Maryland* 27 (May, 5 2001) (unpublished manuscript), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.608.4702&rep=rep1&type=pdf>. Some studies have suggested that white people use and sell drugs at even higher rates than black people. Jonathan Rothwell, *How the War on Drugs Damages Black Social Mobility*, BROOKINGS (Sept. 30, 2014), <https://www.brookings.edu/blog/social-mobility-memos/2014/09/30/how-the-war-on-drugs-damages-black-social-mobility/>; see also Katherine Beckett et al., *Race, Drugs, and Policing: Understanding Disparities in Drug Delivery Arrests*, 44 CRIMINOLOGY 105, 117–18 (2006).

256. Megan Stevenson & Sandra Mayson, *The Scale of Misdemeanor Justice*, 98 B.U. L. REV. 731, 769–70 (2018).

257. However, it is incorrect to draw the conclusion from this that the resulting algorithms are “distorted.” Cf. Barabas et al., *supra* note 255, at 3–4 (stating that using distorted data produces distorted results).

258. Mayson, *supra* note 254, at 2264.

259. See Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15, 16 (2019) (noting that police data may include arrests of people known to be innocent, false claims of criminal activity, and doctored arrest statistics).

people who have served the minimum sentence on the grounds that they might commit a misdemeanor or low-level felony is not justifiable as a normative matter, even if the risk of such offending is high.²⁶⁰ Several recent studies in the pretrial context have found that RAIs so constructed (by using arrests for serious crimes as the outcome measure) do not produce racial bias even in the sense highlighted by the *ProPublica* study.²⁶¹

The second solution to the disparate policing problem—in addition to or instead of changing the outcome measure—is what could be called *race-conscious* calibration of the RAI, which in effect involves creating different algorithms for different ethnicities or races. In Mayson’s New Orleans, a race-conscious RAI might put a black defendant with three arrests in a much lower risk category than a white defendant with three arrests. The possible problem with this approach is that, like classification parity, it violates anticlassification principles; it explicitly takes race into account. But unlike classification parity—which to achieve its version of fairness changes an accurate conclusion about propensity to commit crime²⁶²—race-conscious calibration serves the important state interests of protecting the public and avoiding unnecessary incarceration by rectifying the impact of discriminatory practices that unfairly raise one’s risk score. The classification parity approach nullifies a person’s real risk level. In contrast, race-conscious calibration that corrects for racially biased policing works to nullify a person’s *erroneous* risk level, one that is a consequence of prior government malfeasance and that cannot be justified by any legitimate state interest; thus, it should not be considered a violation of the Equal Protection Clause.²⁶³

260. See ASHWORTH & ZEDNER, *supra* note 51, at 260 (arguing that sentence enhancements based on risk should not occur unless the person “is adjudged to present a very serious danger to others” and the person “has a previous conviction for a very serious offence”).

261. Sarah L. Desmarais et al., *Predictive Validity of Pretrial Risk Assessments: A Systematic Review of the Literature*, 47 CRIM. JUST. & BEHAV. ONLINEFIRST, June 2020, at 17–19; Douglas B. Marlowe et al., *Employing Standardized Risk Assessment in Pretrial Release Decisions: Association with Criminal Justice Outcomes and Racial Equity*, 44 LAW & HUM. BEHAV. 361, 371–72 (2020) (finding no racial disparities in risk prediction but finding such disparities in length of detention and diversion); see also Evan M. Lowder et al., *Improving the Accuracy and Fairness of Pretrial Release Decisions: A Multi-site Study of Risk Assessments Implemented in Four Counties*, NAT’L INST. OF JUST., <https://nij.ojp.gov/funding/awards/2018-r2-cx-0023> (last visited Mar. 15, 2021).

262. See Angwin et al., *supra* note 245.

263. See Aziz Z. Huq, *Constitutional Rights in the Machine Learning State* 31 (U. of Chi. Pub. L. Working Paper, Paper No. 752, 2020) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3613282 (“[A]n official’s mere awareness of race raises no constitutional problem. By analogy, it may also be that mere inclusion of race as a feature of training data should not be per se problematic.”).

Note that this entire discussion about inequality and risk factors is only possible when risk assessment is a product of an RAI. If instead a risk assessment is the product of human judgment, taking any type of ameliorative action would be much more difficult. Even with implicit bias training, humans will find it difficult to avoid the impact of race or race proxies.²⁶⁴ And even with a requirement that reasons for acting be documented, human intuitive predictions of reoffending are opaque, which makes them difficult to challenge.²⁶⁵ In contrast, an RAI displays its stereotyping assumptions on its face. Thus, compared to human-driven risk assessment, RAIs allow bias to be more easily discovered. And because of that transparency, RAIs can take racial biases into account.

For example, using the Post Conviction Risk Assessment tool (a federal probation RAI) and large samples, Jennifer Skeem and Christopher Lowenkamp have shown a number of ways race can be used in constructing a well-calibrated instrument that reduces significantly the “proxy effect” of race.²⁶⁶ Crystal Yang and Will Dobbie were able to do the same thing in the pretrial setting with the PSA, with only minimal impact on predictive validity.²⁶⁷ These sorts of efforts illustrate that fact, as Sendhil Mullainathan asserted in the title to a recent article, *Biased Algorithms Are Easier to Fix than Biased People*.²⁶⁸

2. Sex

While race is not explicitly incorporated into any commonly used RAI, sex—specifically, maleness—is a risk factor in several risk tools.²⁶⁹ In part, that is because developers have realized that an RAI

264. Adam Benforado, *Can Science Save Justice?*, 101 JUDICATURE 24, 26, 28 (2017) (“Studies on sentencing have shown that judges are influenced by the race of the defendant [T]he latest psychological research suggests that much of the skew is not susceptible to conscious control. There is no magic switch to erase a lifetime of exposure to damaging stereotypes that link the concepts of blackness and violence . . .”).

265. *See id.* at 30.

266. Jennifer Skeem & Christopher Lowenkamp, *Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism*, 38 BEHAV. SCI. & L. 259, 263 (2020).

267. CRYSTAL S. YANG & WILL DOBBIE, EQUAL PROTECTION UNDER ALGORITHMS: A NEW STATISTICAL AND LEGAL FRAMEWORK 53–58 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3462379.

268. Sendhil Mullainathan, *Biased Algorithms are Easier to Fix Than Biased People*, N.Y. TIMES (Dec. 6, 2019), <https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html>; *see also* Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 115 (2018) (“Getting the proper regulatory system in place . . . has the potential to turn algorithms into a powerful counterweight to human discrimination and a positive force for social good of multiple kinds.”).

269. *See supra* text accompanying notes 150, 154.

that is well-calibrated for men may not be well-calibrated for women; women do not recidivate as much as men, apparently even when they are otherwise associated with identical risk factors.²⁷⁰ From an empirical point of view, that situation calls for an RAI validated on a female population. But, just as with race, while this sex-conscious classification approach assures better calibration, it may violate the anticlassification principle. For instance, in *Craig v. Boren*,²⁷¹ the Supreme Court struck down a law that allowed women to buy alcohol at age eighteen while prohibiting alcohol sales to males until they were twenty-one, despite evidence that men had higher rates of drunk driving.²⁷² Apparently relying on similar reasoning, statutes in some states specifically ban use of gender in imposing a sentence.²⁷³

Nonetheless, in *State v. Loomis*,²⁷⁴ the Wisconsin Supreme Court suggested that discriminating on the basis of sex is permissible if it validly helps distinguish between males and females in terms of risk.²⁷⁵ Loomis's sentence had been enhanced using the COMPAS, which specifically took gender into account; Loomis argued that this disposition violated due process.²⁷⁶ Although, as a result of this framing, the Wisconsin court did not explicitly address the equal protection issue, it did state, in the course of rejecting Loomis's claim, that "it appears that any risk assessment tool which fails to differentiate between men and women will misclassify both genders."²⁷⁷ In essence, the court was saying that, because of its enhancement to accuracy, incorporating gender was a narrowly tailored means of meeting the state's interest in preventing harm to the public in a cost-efficient manner.

Had the court directly addressed the equal protection issue, would it have had to decide otherwise? *Boren* might be distinguished on the ground that the statistical evidence in that case about the

270. Myrna S. Raeder, *Gender-Related Issues in a Post-Booker Federal Guidelines World*, 37 MCGEORGE L. REV. 691, 694 (2006).

271. 429 U.S. 190 (1976).

272. *Id.* at 192, 210.

273. OHIO REV. CODE ANN. § 2929.11(C) ("A court that imposes a sentence upon an offender for a felony shall not base the sentence upon the race, ethnic background, gender, or religion of the offender."); TENN. CODE ANN. § 40-35-102(4) ("Sentencing should exclude all considerations respecting race, gender, creed, religion, national origin and social status of the individual").

274. 881 N.W.2d 749 (Wis. 2016).

275. *See id.* at 753–54.

276. *Id.* at 753.

277. *Id.* at 766. Other courts have recognized this point. *See, e.g.*, *Karsjens v. Jesson*, 6 F. Supp. 3d 958, 967–68 (D. Minn. 2014) (describing expert testimony in cases challenging female sex offenders' civil commitment programming stating that actuarial risk tools normed on male sex offenders are inapplicable to females); *In re Risk Level Determination of S.S.*, 726 N.W.2d 121, 123 (Minn. Ct. App. 2007) (noting that experts declined to score a sexual recidivism risk tool for a female defendant as it had not been validated on women).

difference between men and women was weak; as Justice Brennan pointed out for the majority, the data showing that men are arrested for drunk driving more often than women might merely mean that “reckless’ young men who drink and drive are transformed into arrest statistics, while their female counterparts are chivalrously escorted home.”²⁷⁸ But in two cases applying Title VII, the federal antidiscrimination statute, the Court held that employers may not require women to pay more into, or receive less from, retirement accounts than men, despite strong data showing that women live longer and thus that, without those adjustments, they will cost the employer more money than men.²⁷⁹ Although Title VII does not apply to the criminal justice system and recognizes a broader scope to discrimination claims, these cases signal the Court’s concern with explicit use of sex as a discriminator.

At the same time, the state’s interest in keeping the female drinking age at eighteen at issue in *Boren*, and even the employer’s interest in saving money in the Title VII cases, pale in comparison to the state’s twin interests that are inevitably at stake in cases like *Loomis*: protecting the public and avoiding unnecessary incarceration (in this case, of women). If Virginia’s NVRA instrument is any guide (recall that, under that instrument, being male brings ten points, being female only one),²⁸⁰ sex is a powerful predictor that can help effectuate both interests. Like race-conscious calibration, sex-conscious calibration makes empirical sense and should not be considered violative of equal protection.

3. Other Traits

Of the factors typically found in RAIs, only race and sex trigger Fourteenth Amendment protection and thus require more than a rational basis for their use under current law. Nonetheless, Sonja Starr has argued that the Fourteenth Amendment also bars RAIs from using poverty or proxies for it (e.g., unemployment, location, or house ownership),²⁸¹ based primarily on the Supreme Court’s pronouncement in *Bearden v. Georgia*²⁸² that revoking parole for an offender who has failed to pay a fine “would be little more than punishing [him] for his poverty” and “would be contrary to the fundamental fairness required by the Fourteenth Amendment.”²⁸³ However, no court has interpreted *Bearden* to mean that factors *related* to poverty are anathema in assessing either risk or

278. *Craig v. Boren*, 429 U.S. 190, 202 n.14 (1976).

279. *Ariz. Governing Comm. for Tax Deferred Annuity & Deferred Comp. Plans v. Norris*, 463 U.S. 1073, 1086 (1983); *City of Los Angeles, Dep’t of Water & Power v. Manhart*, 435 U.S. 702, 711 (1978).

280. *See supra* notes 153–55 and accompanying text.

281. Starr, *supra* note 244, at 830.

282. 461 U.S. 660 (1983).

283. *Id.* at 671, 673.

punishment generally.²⁸⁴ Furthermore, *Bearden* itself stated that “a sentencing court can consider a defendant’s employment history and financial resources in setting an initial punishment”²⁸⁵ and emphasized that the only sentencing practice it was barring was the use of poverty “as the *sole* justification for imprisonment,”²⁸⁶ which no RAI comes close to doing.

This does not mean that any wealth-related risk factor is fair game—the state still must demonstrate a rational basis for its use. The rationale developed above for permitting reliance on risk factors like sex or that correlate with race rested on the assumption that these factors are powerful predictors that help achieve the compelling state interest of protecting society.²⁸⁷ Following equal protection’s tiered analysis, the relevance to risk assessment of other factors—such as age, employment status, home life as a child, diagnosis, or marital status—need not be as robust. At the same time, however, these types of factors should add some nontrivial predictive weight to the algorithm. Because of their minimal predictive value, for instance, employment and marital status were eventually dropped from Virginia’s NVRA.²⁸⁸

In contrast, some RAIs explicitly treat these types of factors as strong predictors of reoffending. For instance, the VRAG includes as risk factors psychopathy, personality disorder, whether the individual’s parents were present in the home at age sixteen, and marital status. Furthermore, these factors, especially the first three, are given significant weight in the algorithm.²⁸⁹ In contrast to some types of arrests, these traits cannot easily be attributed to discrimination, even on the part of society as a whole, much less the government. If one’s intuition remains that it is an injustice to base confinement on such factors even when they clearly improve calibration, it probably stems not from concerns about egalitarian injustice, but rather from worries about what could be called retributive injustice.

284. See, e.g., *United States v. Flowers*, 946 F. Supp. 2d 1295, 1300 (M.D. Ala. 2013) (“[R]elative wealth and poverty will inevitably have some effect on the administration of justice”); *State v. Johnson*, 315 P.3d 1090, 1099 (Wash. 2014) (interpreting *Bearden* to mean that a person cannot be imprisoned for failure to pay a fine).

285. 461 U.S. at 671.

286. *Id.*

287. Huq, *supra* note 263, at 4–5.

288. Email from Meredith Farrar-Owens, Dir., Va. Crim. Sent’g Comm’n, to Christopher Slobogin (March 25, 2020, 9:58 EST) (indicating that these two factors were dropped on July 1, 2013, “based on a study of new felony cases”).

289. See *supra* text accompanying notes 149–50.

B. *The Claim of Retributive Injustice*

While good calibration, adapted to ensure that algorithms are not tainted by state-enabled discrimination, can minimize egalitarian injustice, it may not assuage those who voice a closely related fairness objection, to the effect that many risk factors have nothing to do with individual blameworthiness. Hollywood's images of futuristic societies in which prisoners are selected according to genetic makeup or brain chemistry are far from current reality. But the difference between bio-prediction of this sort and an actuarial score, which often relies on static or congenital factors, is one of degree, not kind.

The retributive injustice claim comes in two forms. The strong form is that risk can never be a legitimate consideration in the criminal process. That claim will not be addressed here, although I have addressed it elsewhere.²⁹⁰ The weaker retributive injustice claim is that even if risk can in whole or part justify criminal punishment, this type of deprivation of liberty must be based solely on culpable conduct.²⁹¹ At its broadest reach, that claim would mean that RAIs cannot rely on conduct that is not criminally blameworthy—such as substance abuse, remaining single, or choosing not to work—nor on circumstances that are not blameworthy in any sense—such as being abandoned by one's parents before majority or living in a particular area. Even more obvious for those who take this perspective, RAIs cannot incorporate traits, such as gender, age, or diagnosis, that have nothing to do with any type of conduct, blameworthy or not. Perhaps the RAI that most dramatically transgresses this point of view is provided by Virginia's NVRA, which, as mentioned above, recommends a prison sentence simply for being under twenty-one, being male, and having one prior conviction.²⁹²

In *Buck v. Davis*, the Supreme Court appeared to recognize the retributive injustice concern. As noted earlier, the thrust of *Davis* was a rebuke of egalitarian injustice; the Court stated that using the fact of being black as a risk factor “appealed to a powerful racial stereotype” and “coincided precisely with a particularly noxious strain of racial prejudice”²⁹³ But the Court went on to say that sentencing a person on the basis of race “is a disturbing departure from a basic premise of our criminal justice system: Our law punishes

290. See Christopher Slobogin, *A Defence of Modern Risk-Based Sentencing*, in *PREDICTIVE SENTENCING: NORMATIVE AND EMPIRICAL PERSPECTIVES* 107, 116–21 (Jan Willem de Keijser et al., eds., 2019) (Hart Publishing).

291. Michael Tonry, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 *FED. SENT'G REP.* 167, 169 (2014) (“Ascribed characteristics for which individuals bear no responsibility, such as race, ethnicity, gender, and age, should not be included . . .”).

292. See Kern & Owens, *supra* note 153, at 166.

293. *Buck v. Davis*, 137 S. Ct. 759, 776 (2017).

people for what they do, not who they are.”²⁹⁴ Taken literally, this latter sentiment would prohibit punishers from relying not only on race but also on risk factors such as gender and age, and probably on factors involving current mental state, such as diagnosis or lack of insight, since none of these variables involve conduct.

At the same time, the Supreme Court on several occasions has explicitly permitted death sentences to be imposed on the ground that the offender is “dangerous,”²⁹⁵ including in *Barefoot v. Estelle*,²⁹⁶ where the state’s expert opinion about dangerousness was predicated on a diagnosis.²⁹⁷ Furthermore, the provenance for the Court’s declaration in *Davis* that people cannot be punished for “who they are” is not at all clear: while *Davis*’s ban on race presumably comes from the Equal Protection Clause, the opinion did not identify any particular constitutional provision when it made its more general pronouncement about the proper basis of punishment.²⁹⁸ The most likely candidate is the Eighth Amendment’s ban on cruel and unusual punishment, which has been construed to forbid criminalizing status, specifically the status of being addicted.²⁹⁹ But the people who are subjected to risk assessment at sentencing have already been convicted for criminal *conduct*, defined by statutes that are presumably constitutional. And the courts, including the Supreme Court, have long permitted sentences to be based on risk.³⁰⁰ So in the end, the *Davis* decision is most accurately described as a prohibition on the use of race in sentencing, rather than as a wholesale rejection of incarceration based on the status of being high risk.

294. *Id.* at 778.

295. *Jurek v. Texas*, 428 U.S. 262, 275–76 (1976) (stating that “prediction of future criminal conduct is an essential element in many of the decisions rendered throughout our criminal justice system [mentioning bail, sentencing, and parole determinations as examples of such decisions]. . . . The task that a Texas jury must perform in answering the statutory question [about dangerousness of a capital defendant] is thus basically no different from the task performed countless times each day throughout the American system of criminal justice.”).

296. 463 U.S. 880, 884 (1983).

297. *Id.* at 919 (Blackmun, J., dissenting) (“[Dr. Grigson] placed Barefoot in the ‘most severe category’ of sociopaths (on a scale of one to ten, Barefoot was ‘above ten’), and stated that there was no known cure for the condition.”).

298. *Davis*, 137 S. Ct. at 778.

299. *Robinson v. California*, 370 U.S. 660, 666–67 (1962) (holding that criminalizing the status of being a narcotics addict violates the Eighth Amendment).

300. *United States v. Salerno*, 481 U.S. 739, 748–49 (1987) (allowing pretrial detention to be based on risk); *Pennsylvania ex rel. Sullivan v. Ashe*, 302 U.S. 51, 55 (1937) (noting that the government “may inflict a deserved penalty merely to vindicate the law or to deter or to reform the offender or for all of these purposes. . . . [The offender’s] past may be taken to indicate his present purposes and tendencies and significantly to suggest the period of restraint and the kind of discipline that ought to be imposed upon him.”).

While *Davis* may not squarely pose a retributive injustice challenge to RAIs, some scholars have done so, arguing that a criminal justice system unmoored from just desert is immoral because of its insult to autonomy and dignity. The point was put succinctly by Andrew von Hirsch, a noted retributivist, who posited that “[u]nless the person actually made the wrongful choice he was predicted to make, he ought not to be condemned for that choice—and hence should not suffer punishment for it.”³⁰¹ That view might permit risk-oriented punishment based on prior crimes as well as the current crime. But, as applied to RAIs, it would prohibit not only risk factors that do not consist of conduct but also risk factors based on conduct that is not criminally blameworthy, such as choices about psychoactive substance use, employment, or education. Accordingly, it would put an end to modern risk assessment.

From a risk assessment perspective, there are two significant practical problems with limiting risk factors to criminal conduct: First, removal of all non-crime factors from an RAI is likely to substantially reduce accuracy. Second, restricting RAIs to crime-only risk factors is also likely to create egalitarian injustice. A young male with psychopathic tendencies and one prior crime represents a much higher risk than an older female suffering from anxiety who has committed the same crime; yet, under a crime-only approach, both would be treated identically in terms of risk.

A true retributivist worried about risk’s insult to blameworthiness would not be dissuaded by these types of concerns. Rather, more conceptual counterarguments are required. A first such response to the retributive injustice claim is that it is based on a category mistake. Interventions based on risk are not about culpability for past conduct; they are focused on preventing future conduct. Culpability assessments are the province of trial and, as discussed in Part II, they should also be crucial in setting the range of sentencing options.³⁰² But if risk is a legitimate consideration at the post-trial stage, it is a separate inquiry from blameworthiness and should not be entangled with it.

The retributive inquiry and the risk inquiry are usually orthogonal to one another. While youth is often considered a mitigating factor from a retributive point of view, it is clearly an aggravating factor from a risk perspective.³⁰³ The fact that one was abandoned by one’s parents at age sixteen might be considered a

301. ANDREW VON HIRSCH, PAST OR FUTURE CRIMES: DESERVEDNESS AND DANGEROUSNESS IN THE SENTENCING OF CRIMINALS 11 (1985).

302. See *supra* Subpart II.A.

303. Compare SENTENCING MULTIPLE CRIMES 258 (Jesper Ryberg, Julian V. Roberts & Jan W. de Keijser, eds., 2017) (considering youth a mitigating factor), with Elizabeth F. Emens, *Aggravating Youth: Roper v. Simmons and Age Discrimination*, 1 SUP. CT. REV. 51, 75–79 (2005) (considering youth an aggravating factor).

mitigator when focused on culpability, but to the developers of the VRAG, it is a risk factor.³⁰⁴ Or consider criminal history itself. It may be an aggravator from both a blameworthiness and a risk perspective, but rarely in precisely the same way. A third property crime might be considered highly blameworthy nose-thumbing to a retributivist, but not particularly indicative of high risk; alternatively, a retributivist might consider the two previous convictions irrelevant once the associated sentences have been served, but if risk assessment is the goal, these crimes presumably would be highly probative of future conduct.³⁰⁵ Risk and culpability are clearly conceptually separate inquiries.

The first two examples also put the lie to the implicit claim by those who make the retributive injustice argument that culpability assessments are never based on status. Youth is a status. Yet, it is routinely treated as a mitigator by retributivists;³⁰⁶ indeed, in his latest work, von Hirsch himself has proposed a “youth discount” for punishment and argued that it should be “categorical,” not individualized, which means that simply being youthful would require leniency.³⁰⁷ Likewise, parental abandonment is a status. But a retributivist might treat it as a mitigator.

One cannot avoid this contradiction by claiming that, as long as noncriminal factors are used only as a mitigator, the damage to dignity and autonomy is minimized. If people are provided leniency based in part on their youth, older people are being treated more harshly because of their status. If people who were left by their parents are given a break, people with intact families are not. Status permeates both risk-based and retribution-based regimes.

At the same time, contrary to the literal interpretation of the Supreme Court’s decision in *Buck v. Davis*, a person’s risk categorization is not simply a status, but rather is closely associated with blameworthy choices, in two ways. First, risk assessment is, in large part, an evaluation of what a person has chosen to do. That is because it is, in essence, an evaluation of one’s character, which on many accounts, including from scholars such as Peter Arenella,

304. Harris et al., *supra* note 149, at 382, 387.

305. The debate among retributivists as to how to treat punishment for multiple crimes has produced numerous positions which are not reconcilable with one another. *See generally* SENTENCING MULTIPLE CRIMES, *supra* note 303.

306. *See, e.g.*, *Roper v. Simmons*, 543 U.S. 551, 571–73 (2005) (quoting *Atkins v. Virginia*, 536 U.S. 304, 319 (2002)) (“Once the diminished culpability of juveniles is recognized, it is evident that the penological justifications for the death penalty apply to them with lesser force than to adults,” and rejecting the argument that it is “arbitrary and unnecessary to adopt a categorical rule barring imposition of the death penalty on any offender under 18 years of age”).

307. ANDREW VON HIRSCH, *DESERVED CRIMINAL SENTENCES* § 11.2c (2017).

James Whitman, and Kyron Huigens, is directly relevant to desert.³⁰⁸ The Supreme Court itself has made the connection when it stated in *Deck v. Missouri*³⁰⁹ that “character and propensities of the defendant are part of a ‘unique, individualized judgment regarding the punishment that a particular person *deserves*.’”³¹⁰ Character is an amalgam of choices—choices that are often constrained by circumstances, but choices nonetheless. People make decisions not only about whether to engage in antisocial conduct but about their friends, family life, education, work, the places they frequent, the amount of drugs or alcohol they ingest, and whether to seek treatment for emotional problems such as anger and impulsivity—all of which are examples of precisely the types of activities captured in the most sophisticated RAIs.³¹¹ If these choices combine to make one’s character high risk, they could be said to be blameworthy, even if they do not involve criminal activity.³¹²

Of course, character analysis alone does not justify using risk factors that have nothing to do with choice, such as age, gender, and diagnosis. But the ultimate, and best, argument against the retributive injustice claim is that viewed properly, risk factors—whether or not they are the product of choice—are not the *reason* for

308. Peter Arenella, *Character, Choice and Moral Agency: The Relevance of Character to Our Moral Culpability Judgments*, 7 SOC. PHIL. & POL’Y 59, 61 (1990) (arguing that there is no means of judging persons except through assessing their character); Kyron Huigens, *The Dead End of Deterrence, and Beyond*, 41 WM. & MARY L. REV. 943, 1022–34 (2000) (describing the aretaic theory of punishment, which aims at improving character); James Q. Whitman, *The Case for Penal Modernism: Beyond Utility and Desert*, 1:2 CRITICAL ANALYSIS LAW 143, 178 (2014) (arguing for a European style trial mixing assessment of desert and character, because such a trial “makes it possible to consider the full spectrum of information about individual blameworthiness, including both dangerousness and deservingness”).

309. 544 U.S. 622 (2005).

310. *Id.* at 633 (emphasis added) (quoting *Zant v. Stephens*, 462 U.S. 862, 900 (1983) (Rehnquist, J., concurring)). Making the link between risk and character even more explicit, just prior to this statement, the Court stated that “danger to the community . . . almost inevitably affects adversely the jury’s perception of the defendant’s character.” *Id.* at 622–23. The Court has even adhered to this sentiment in cases involving young offenders, whose character is typically in its formative stages; thus, for instance, while emphasizing such situations should be rare, in *Graham v. Florida*, the Court stated: “Those who commit truly horrifying crimes as juveniles may turn out to be irredeemable, and thus *deserving* of incarceration for the duration of their lives.” 560 U.S. 48, 75 (2010) (emphasis added).

311. *See supra* Subpart III.A.

312. *Cf.* Douglas Husak, *Lifting the Cloak: Preventive Detention as Punishment*, 48 SAN DIEGO L. REV. 1173, 1195 (2011) (“As long as the possession of the characteristic *x*, *y*, and *z* are under the control of persons the state preventively detains—as I would insist—punishment would be compatible with the principle I propose to substitute for the act requirement.”).

an offender's sentence. Rather, they are merely *evidence* of what a person will decide to do in the future, in the same way that a finding of a blameworthy choice in the past may rely on various facts that are not culpable in themselves, such as marriage to the victim, presence near the scene of the crime, or possession of a weapon. Risk assessments try to predict future culpable choices, just as circumstantial evidence about *actus reus* and *mens rea* tries to postdict culpable choices. Consider again the Virginia instrument that puts so much weight on age and gender.³¹³ While it is extremely unlikely that those two variables alone provide sufficient predictive power, if they did and a person was imprisoned rather than diverted to the community as a result, it would not be *because* of youth and gender. Rather, it would be because the RAI indicates he belongs to a group of people who pose a high probability of choosing to reoffend.

Risk is not pristine desert, but it is not some soulless mechanical assessment of humans-as-machines either. The debate on this score has been more hyperbolic than productive.

C. *The Claim of Procedural Injustice*

The fact that risk assessment does not consist simply of converting people into probability numbers or abstract categories such as high or low risk does not mean that it is not perceived that way. Underlying much of the Bill of Rights is the notion that according dignity to people suspected or convicted of crime is intrinsically valuable. A suspect may not be subject to search or seizure without strong justification, those who are accused have the rights to counsel, public trial, and confrontation, and punishment cannot be cruel and unusual, but rather must be consistent with “evolving standards of decency”³¹⁴ Additionally, the procedural justice literature suggests that for a legal process to be considered legitimate—especially by those enmeshed in it—it must treat people with respect, by giving them voice, ensuring a transparent process, and providing explanations for any decisions made.³¹⁵ Adhering to

313. See *supra* notes 152–54 and accompanying text.

314. *Trop v. Dulles*, 356 U.S. 86, 101 (1958) (interpreting the scope of the Eighth Amendment's cruel and unusual punishment clause). See also Judith Resnik & Julie Chi-hye Suk, *Adding Insult to Injury: Questioning the Role of Dignity in Conceptions of Sovereignty*, 55 STAN. L. REV. 1921, 1935 (2003) (“[W]e find dignity mentioned in relation to the Eighth Amendment prohibition on cruel and unusual punishment; the Fourth Amendment protection against unlawful searches and seizures; the Fourteenth and Fifteenth Amendment rights to be free from discrimination, and the Ninth and Fourteenth Amendment rights to make one's own decisions on procreation.”).

315. The foundational research comes from TOM R. TYLER, *WHY PEOPLE OBEY THE LAW* 115–17 (1990) (describing studies that suggest that perceptions of fairness hinge on whether disputants feel they have been given a voice in the

those goals can promote compliance with legal decisions and a more general sense of an obligation to obey the law, whereas a failure to do so may lead to less cooperation, not only with the particular decision but with the government that permits it to happen.³¹⁶

Procedural injustice could be a real danger in an RAI-oriented regime unless certain precautions are taken.³¹⁷ While, compared to RAIs, clinical predictions may be more susceptible to bias, less accurate, and equally influenced by stereotypes, they *appear* to be more individualized, especially when based on an interview and framed in psychological terms. That apparent difference may be exacerbated if, as this Article recommends, RAI results are given presumptive effect on the issue of risk. Thus, procedures designed to ensure that litigants have a voice in the process and can challenge the internal workings of the RAI results are essential.

1. Voice

If, as Part III argued, postconviction proceedings required counsel, a truly adversarial process, and transparent algorithms (the latter issue addressed in more detail below), concerns about procedural injustice would be substantially diminished. RAIs could be challenged in several ways. First, offenders would be able to present their own RAI results. Second, the defense could attack the accuracy of a conclusion that a particular risk factor is present (e.g., the validity of an assumed arrest or conviction, the applicability of a diagnosis, or the failure to complete a program that in fact was not available to the offender). Third, defendants could proffer protective factors that were not considered by the developers of the instrument (e.g., completion of a treatment or educational program, changes in employment status); researchers are beginning to identify a number of such factors.³¹⁸ The latter type of evidence should be particularly

process and are treated with dignity, as well as whether outcomes are fair over time).

316. *Id.* at 56 (suggesting that people comply with the law for a complex set of reasons that include cost-benefit analysis, the norms of peers, one's own norms, and the perceived legitimacy of the authorities, but concluding that the last reason is the most important for policymakers, because titrating deterrence and changing individual and group norms are both difficult); *see also* Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUST. 283, 297 (2003) (discussing a study that suggests that "procedural justice" enhances the public's view surrounding the legitimacy of rules and authority and makes individuals feel obligated to follow the law).

317. *See* Michael O'Hear, *Actuarial Risk Assessment at Sentencing: Potential Consequences for Mass Incarceration and Legitimacy*, 38 BEHAV. SCI. & L. 193, 195–96 (2020).

318. *See* Richard B. A. Coupland & Mark E. Olver, *Assessing Protective Factors in Treated Violent Offenders: Associations with Recidivism Reduction and Positive Community Outcomes*, in 32 PSYCH. ASSESSMENT: A J. OF CONSULTING & CLINICAL PSYCH. 493, 494 (2020).

useful in humanizing and individualizing the process. It would also help combat adversarial bias, which research has shown afflicts even testimony based on relatively objective RAIs.³¹⁹

Two other procedural components are crucial, especially if the arguments advanced in Part II for a more formal process do not succeed.³²⁰ First, evaluators, judges, and parole boards must lay out for defendants how the RAI works and why it reached the conclusions about risk it did. RAIs can tell a story similar to one that a clinical expert would tell. This is true even if risk factors are primarily static. While such factors usually only correlate with risk rather than explain it, theories do exist as to why they are relevant to predictive validity (e.g., youth are more impulsive and subject to peer pressure; people with poor childhoods have fewer adaptive skills; people with criminal histories tend to repeat them).³²¹ Explanation along these lines can meaningfully diminish the perception of robot justice. For instance, one study of the procedural justice implications of risk algorithms found that while lay subjects preferred clinical to actuarial judgment in the abstract, their preferences were reversed when they were informed that the algorithm was more accurate, and they were even more likely to prefer algorithms when the factors used to construct them were made transparent.³²²

At the same time, RAIs ideally should contain dynamic factors as well, because these signal that the individual has some control over his or her fate. Evaluators and decision-makers should make sure to describe these factors, particularly protective ones. Likewise, if the determination is made that intervention is necessary because of the person's risk, procedural justice would be enhanced if the decision-maker specifies the types of actions the individual can take to reduce that risk (e.g., substance abuse treatment, cognitive therapy, employment). Some RAIs, such as the HCR-20, are more attuned to this goal because they include dynamic or variable factors that the defendant can do something about.³²³ This information can communicate to the individual that, whatever the numbers may say, he or she has the ability to change them.

A second crucial procedural component, whether or not counsel is involved, is ensuring that RAIs are subject to legislative and administrative review. For instance, the PATTERN is the result of the First Step Act of 2018, in which Congress directed that (1) the

319. Stephane M. Shepherd & Danny Sullivan, *Covert and Implicit Influences on the Interpretation of Violence Risk Instruments*, 24 *PSYCHIATRY PSYCH. & L.* 292, 297 (2017) (citing studies).

320. See discussion *supra* Subpart II.C.

321. See R. Karl Hanson, *Giving Meaning to Risk Factors*, 15 *PUB. SAFETY CANADA*, Nov. 2010, at 1, 1.

322. A.J. Wang, *Procedural Justice and Risk Assessment Algorithms* 20 (June 21, 2018) (unpublished manuscript) (on file with author).

323. Douglas & Webster, *supra* note 161, at 8.

Attorney General develop and release a “risk and needs assessment system” to determine the “recidivism risk of each prisoner” following “an objective and statistically validated method,” (2) a panel of researchers approve the instrument, (3) the instrument be annually validated, and (4) the Bureau of Prison staff “demonstrate competence in administering the [PATTERN], including interrater reliability, on a biannual basis.”³²⁴ Given that such an instrument is in effect an informal rule governing “the rights and obligations” of citizens, such a system could be subject to notice and comment requirements under the Administrative Procedure Act;³²⁵ in fact, the Federal Bureau of Prisons (“BOP”) did call for public comment and responded to criticisms about both the validity of the PATTERN instrument and its relevance and fairness.³²⁶ Ideally, any instrument produced through such a process would be subject to hard-look review to ensure that the government can demonstrate it is rationally related to the state’s objectives.³²⁷ Where those instruments are produced by private companies, Andrea Nishi has made the argument that the private nondelegation doctrine—the “lesser-known cousin” of the doctrine applicable to government agencies³²⁸—applies in this setting and requires that “grants of government power to private entities are adequately structured to preserve constitutional accountability.”³²⁹

This upper-level review of RAIs is essential. A federal panel such as that created by the First Step Act is one possibility; the now-defunct National Commission of Forensic Science would have been

324. 18 U.S.C. §§ 3632(a)(1), 3632(f)(4), 3635(6).

325. *Long Island Care at Home, Ltd. v. Coke*, 551 U.S. 158, 172–73 (2007) (associating notice-and-comment rulemaking with a regulation that “directly governs the conduct of members of the public, ‘affecting individual rights and obligations’” (quoting *Chrysler Corp. v. Brown*, 441 U.S. 281, 302 (1979))); *cf.* Richard A. Bierschbach & Stephanos Bibas, *Notice-and-Comment Sentencing*, 97 MINN. L. REV. 1, 25, 32 (2012) (comparing the effects of criminal sentencing guidelines on citizens to the rationale behind notice and comment procedures).

326. I was one of those who testified on behalf of the American Bar Association’s Criminal Justice Section, consistent with the points made in this Article. A summary of the responses to the notice and comment period can be found at The First Step Act Risk and Needs Assessment System. U.S. DEPT OF JUST., *supra* note 32, at 25–37. The Department of Justice’s Internal Review Committee did not accept all of the suggestions but is committed to continually receiving feedback and revalidation efforts. *Id.* at 84–85.

327. Kevin M. Stack, *Interpreting Regulations*, 111 MICH. L. REV. 355, 379 (2012) (“Hard-look review . . . has long been understood as requiring a higher standard of rationality than the minimum rational basis standard of constitutional review.”).

328. *Ass’n of Am. R.Rs. v. U.S. Dep’t of Transp. (Amtrak I)*, 721 F.3d 666, 670 (D.C. Cir. 2013), *vacated*, 575 U.S. 43 (2015).

329. Andrea Nishi, *Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment*, 119 COLUM. L. REV. 1671, 1695 (2019) (quoting Gillian E. Metzger, *Privatization as Delegation*, 103 COLUM. L. REV. 1367, 1456 (2003)).

another.³³⁰ European countries have forensic institutes that are established to investigate science used in the courts; for instance, in the United Kingdom, the Ministry of Justice houses the Correctional Services Accreditation and Advice Panel, which evaluates a tool in terms of whether it “does what it aims to do.”³³¹ However, for empirical reasons having to do with assuring the RAI is valid for the local population, situating validation responsibility at the state level and ensuring that the state entity takes into account significant jurisdictional differences (particularly urban-rural divides) makes the most sense.³³² In fact, a number of state legislatures have mandated that sentencing judges and corrections officials use a “validated risk assessment tool,”³³³ and in other states, the state sentencing commission,³³⁴ the department of corrections,³³⁵ the state courts generally,³³⁶ or—as California has done with respect to pretrial risk assessments—the courts in each jurisdiction,³³⁷ have taken on the task. As long as the entity employs or can pay for experts who can develop and validate RAIs and ensure that the validation is peer-reviewed by independent experts, it could fulfill a vital procedural role in a risk assessment regime.³³⁸

330. See Jules Epstein, *The National Commission on Forensic Science: Impactful or Ineffectual?*, 48 SETON HALL L. REV. 743, 748–49 (2018) (describing the short-lived commission that was disbanded in 2017 by then-Attorney General Jeff Sessions).

331. Her Majesty’s Prison & Prob. Ser., *Guidance: Risk Assessment of Offenders*, GOV’T DIGIT. SERV. (May 15, 2019), <https://www.gov.uk/guidance/risk-assessment-of-offenders#what-makes-a-good-risk-assessment-tool>.

332. See Alicia Solow-Niederman et al., *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L.J. 705, 724–40 (2019) (noting that the effect of proxies, Simpson’s paradox, and thresholding decisions need to be considered in designing RAIs to be used in diverse jurisdictions).

333. See, e.g., KY. REV. STAT. ANN. § 439.335(1) (West) (“In considering the granting of parole and the terms of parole, the parole board shall use the results from an inmate’s validated risk and needs assessment and any other scientific means for personality analysis that may hereafter be developed.”).

334. See, e.g., KAN. SENT’G COMM’N, JUSTICE REINVESTMENT INITIATIVE IN KANSAS 3 (2015), https://www.sentencing.ks.gov/docs/default-source/publications-reports-and-presentations/ksc_jri_report.pdf?sfvrsn=2; UTAH SENT’G COMM’N, 2017 ADULT SENTENCING AND RELEASE GUIDELINES 6 (2020), <https://justice.utah.gov/wp-content/uploads/2020-Adult-Sentencing-and-Release-Guidelines.pdf>.

335. See, e.g., WASH. REV. CODE ANN. § 9.94A.500 (West) (2020).

336. See, e.g., Ariz. Jud. Branch, *Evidence-Based Practice*, www.azcourts.gov/apsd/Evidence-Based-Practice/PresentenceReport (last visited Mar. 15, 2021).

337. PRETRIAL DET. REFORM WORKGROUP, PRETRIAL DETENTION REFORM: RECOMMENDATIONS TO THE CHIEF JUSTICE 52–53 (2017), <http://www.courts.ca.gov/documents/PDRReport-20171023.pdf>.

338. Serious consideration should also be given to including non-experts on the RAI review panel. This is the suggestion of Ngozi Okidegbe in *The Democratizing Potential of Algorithms?* 53 U. CONN. L. REV. (forthcoming 2021).

2. Transparency

None of this is possible, however, if the risk algorithm is not made available for evaluation. Recall that the company that produces the COMPAS refuses to reveal its algorithm or the weights assigned to risk factors, claiming trade secret protection—a claim that the Wisconsin Supreme Court upheld.³³⁹ Thus, the inner workings of the instrument are hidden. For instance, sophisticated reverse-engineering, well beyond the pay grade of defense attorneys or judges, is needed to figure out the fact that while the COMPAS contains more than one hundred factors, over half of the risk score it produces is attributable to a single factor: the offender's age.³⁴⁰

Even purportedly publicly developed instruments can be less than transparent. Congress required that the PATTERN be made public,³⁴¹ but did not require that the validation procedure that led to development of the instrument nor the data underlying it be disclosed. When asked for more information, the authors of the instrument stated that state-law-driven privacy concerns prevented release even of anonymized versions of the data to outside researchers.³⁴² A number of states have responded to similar requests in the same fashion.³⁴³

The integration of sophisticated machine learning into RAI construction could make matters worse, since under some versions of that technique the weights assigned to risk factors and even the

(“[T]he exclusion of these communities within algorithmic governance operates to reinforce and legitimize the barriers that already impede these communities’ ability to challenge or gain control over the very criminal justice institutions responsible for their oversurveillance, overcriminalization, and overincarceration.”).

339. *State v. Loomis*, 881 N.W.2d 749, 761 (Wis. 2016) (holding that because Loomis had access to the questions the COMPAS asked and the risk assessment itself, he did not need access to “how the risk scores are determined or how the factors are weighed”).

340. Megan T. Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 WASH. U. L. REV. 681, 688–98 (2018).

341. 18 U.S.C. § 3631(b)(4) (requiring the Attorney General, *inter alia*, to “on an annual basis, review, validate, and release publicly on the Department of Justice website the risk and needs assessment system”).

342. U.S. DEP’T OF JUST., THE FIRST STEP ACT OF 2018: RISK AND NEEDS ASSESSMENT SYSTEM—UPDATE 14–15 (Jan. 2020), <https://www.bop.gov/inmates/fsa/docs/the-first-step-act-of-2018-risk-and-needs-assessment-system-updated.pdf>.

343. Nicholas Diakopoulos, *We Need to Know the Algorithms the Government Uses to Make Important Decisions About Us* (May 23, 2016, 8:48 PM), <https://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869> (noting that only one state responded fully to Freedom of Information Act requests for algorithm documents and source codes).

identity of those factors are inaccessible to humans.³⁴⁴ Furthermore, even if the black box can be opened, serious interpretation problems can arise. More specifically, as Andrew Selbst and Solon Barocas note, some versions of machine learning can be either “inscrutable”—meaning that even when a model is available for direct inspection it may “defy understanding”—or “non-intuitive”—meaning that even where a model is understandable it may “rest on apparent statistical relationships that defy intuition.”³⁴⁵

Proprietary interests and algorithmic opacity could stymie meaningful challenges to RAI results and empirical investigations of its validity. Egalitarian and retributive justice cannot be evaluated without knowing whether risk scores are based on race, gender, age, wealth classifications, or proxies for them and the extent to which they purport to help the state achieve its aim in evaluating risk. The accuracy of the probabilities and other results reached by an RAI cannot be confirmed unless the underlying data and the empirical analysis using it can be evaluated by others. And decision-makers cannot know whether to permit or engage in “adjustments” to a risk assessment based on factors not considered in the instrument unless they know what those factors are.

Once again, Supreme Court jurisprudence provides grounds for contesting this situation. In *Gardner v. Florida*,³⁴⁶ the defendant argued that, before his death sentence was imposed, he had a due process right to discover and rebut the contents of his presentence report.³⁴⁷ The government objected to this claim on a number of grounds: such discovery, it claimed, would make sources reluctant to provide information, would delay the process, might disrupt rehabilitation given the psychological information contained in such reports, and was not necessary given the expertise of judges.³⁴⁸ But the Supreme Court rejected all of these arguments (many of which might be made in defending informal use of RAIs as well), stating: “Our belief that debate between adversaries is often essential to the truth-seeking function of trials requires us also to recognize the importance of giving counsel an opportunity to comment on facts which may influence the sentencing decision”³⁴⁹

344. See generally Doaa Abu Elyounes, *Bail or Jail? Judicial Versus Algorithmic Decision-Making in the Pretrial System*, 21 COLUM. SCI. & TECH. L. REV. 376 (2020) (describing various types of machine learning algorithms and possible difficulties with discerning how they work, but also noting that most RAIs today rely on “traditional regression analysis” and are transparent with respect to the factors considered and the weight they are given).

345. Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1091 (2018).

346. 430 U.S. 349 (1977).

347. *Id.* at 353–55.

348. *Id.* at 358–60.

349. *Id.* at 360.

Gardner involved capital punishment, where the Supreme Court has been particularly meticulous about accuracy. But in *Roviaro v. United States*,³⁵⁰ involving a simple drug case, the Court similarly held that the identity of confidential informants must be revealed to the defendant when the informant possesses facts that are relevant to the defense.³⁵¹ Although it involved a confidential informant rather than a confidential algorithm, *Roviaro* establishes that even strong claims of a need for secrecy (here protecting an informant) should not prevail when the information is crucial to the case.³⁵² While *Roviaro* has been given short shrift in more recent lower court decisions,³⁵³ its central rationale has not been abandoned.³⁵⁴ Some lower courts have followed the logic of these opinions in requiring that defendants be given the facts and opinions underlying their proposed sentences and an opportunity to rebut them.³⁵⁵

Scholars have also made subconstitutional arguments in favor of open algorithms. Danielle Citron has contended that private companies that seek public money for products that affect public policy should not be able to hide behind trade secret laws,³⁵⁶ and Rebecca Wexler has noted that companies' concerns about giving competitors an advantage or discouraging innovation are overblown, especially if protective orders or in camera review requirements are imposed.³⁵⁷ Special attention has been paid to the opacity problems created by machine learning. Most prominently, scholars have

350. 353 U.S. 53 (1957).

351. *Id.* at 64–65.

352. See Zathrina Zasell Gutierrez Perez, Note, *Piercing the Veil of Informant Confidentiality: The Role of In Camera Hearings in the Roviaro Determination*, 46 AM. CRIM. L. REV. 179, 192–94 (2009) (discussing the *Roviaro* holding).

353. See *id.* at 201–13 (describing United States Circuit Courts of Appeal approaches to *Roviaro*).

354. *Id.*

355. *United States v. Millán-Isaac*, 749 F.3d 57, 70 (1st Cir. 2014); *Smith v. Woods*, 505 F. App'x 560, 568 (6th Cir. 2012); *United States v. Hayes*, 171 F.3d 389, 394 (6th Cir. 1999).

356. Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1290–91 (2008) (“[T]he public . . . [and] government actors are unable to influence policy when it is shrouded in closed code.”).

357. See Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1403–13 (2018) (arguing that, given protective orders and other procedural devices, trade secret privilege is unnecessary in criminal cases); see also Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 26 (2014) (“There is little evidence that the inability to keep such systems secret would diminish innovation.”). It is noteworthy that intellectual property claims have also been rejected in civil cases when the potential for error resulting from opacity is “obvious” and “substantial.” See *K.W. v. Armstrong*, 180 F. Supp. 3d 703, 716–17 (D. Idaho 2016).

argued for a “right to explanation,”³⁵⁸ a right that the European Union has explicitly recognized in its General Data Privacy Regulation.³⁵⁹

That right is particularly important in the criminal context, for the reasons advanced in *Gardner*. Even if it turns out that advanced RAIs are demonstrably more accurate than simpler versions (which is unlikely),³⁶⁰ they should be banned from criminal proceedings, at least when they are “inscrutable;” litigants, policymakers, and decision-makers must be provided with understandable information about how they work.³⁶¹ Specifically, developers ought to provide:

[A] complete description of the design and testing process . . . , [a] list of factors that the tool uses and how it weighs them, [t]he thresholds and data used to determine labels for risk scores, . . . [t]he outcome data used to develop and validate the tool at an aggregate and privacy-protecting level, disclosing breakdown of rearrests by charge, severity of charge, . . . age, race, and gender [and] clear definitions of what an instrument forecasts and for what time period.³⁶²

Without the ability to investigate the basis of decisions about risk, injustice—not only procedural injustice but also egalitarian and retributive injustice—will have fertile ground in which to flourish.

358. See Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 209–17 (2019) (reviewing the literature).

359. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. (L 119) 40–43 (EU).

360. ALEXANDRA CHOULDECHOVA & KRISTIAN LUM, *THE PRESENT AND FUTURE OF AI IN PRE-TRIAL RISK ASSESSMENT INSTRUMENTS 3* (2020) (“AI technologies are not likely to achieve considerably greater predictive accuracy than currently available risk assessment instruments.”); Desmarais & Zottola, *supra* note 36, at 813–16 (surveying research and concluding that “these findings suggest that there is no real advantage to using complex statistical models that are challenging for the layperson to understand”).

361. Selbst & Barocas, *supra* note 345, at 1110 (“[With respect to providing transparency, r]esearchers have developed at least three different ways to respond to the demand for explanations: (1) purposefully orchestrating the machine learning process such that the resulting model is interpretable; (2) applying special techniques after model creation to approximate the model in a more readily intelligible form or identify features that are most salient for specific decisions; and (3) providing tools that allow people to interact with the model and get a sense of its operation.”).

362. DAVID G. ROBINSON & LOGAN KOEPKE, *CIVIL RIGHTS AND PRETRIAL RISK ASSESSMENT INSTRUMENTS 11* (2019); see also Kleinberg et al., *supra* note 268, at 2 (“[A]ll the components of an algorithm (including the training data) must be stored and made available for examination and experimentation.”).

V. CONCLUSION—THE NEED FOR EXPERIMENTATION

Risk and needs assessment instruments are crucial tools for pinpointing the hundreds of thousands of offenders who can, with relative safety, be diverted to community programs or be released with no restrictions. Without the quantitative clarity and authority of these instruments, governments will have neither the wherewithal nor the will to make serious inroads on our incarcerated populations. Given retributive urges, decriminalization will at most affect the lowest-level misdemeanors. For the same reason, significant reductions in sentences for more serious crimes are unlikely to be countenanced by the American public unless those reductions take place on an individualized basis and can be shown to be of lower-risk offenders. And without concrete proof that particular offenders are low risk, our elected and politically appointed decision-makers are, understandably, unlikely to opt against confinement.

In short, if the goal is to make significant inroads on the incarcerated population in the United States, risk assessment technology may be the only realistic method of doing so. RAIs would have an even greater impact if they are given presumptive effect. And they should have greater impact still if they are used within a system of preventive justice that relies on risk to calibrate the nature and length of sentences within a retributive framework.

Many jurisdictions are already using RAIs. But very few vet those instruments through a peer review process, give their results presumptive effect, or train judges, lawyers, and correctional officials in their use. Until a few pioneering jurisdictions give RAIs a fair shot, we cannot know whether the hypotheses that this Article advances about their benefits will be borne out. If the integration of well-constructed and presumptively applied RAIs into postconviction settings does not significantly reduce incarcerated populations, recidivism rates, and prison costs, it is probably not worth pursuing on a nationwide basis. But the bet here is that, without such experimentation, the American carceral system will be stuck where it is now for some time to come.