

# REPROGRAMMING FAIRNESS: AFFIRMATIVE ACTION IN ALGORITHMIC CRIMINAL SENTENCING

Jacob D. Humerick\*

## ABSTRACT

When sentencing a convicted criminal defendant, judges in many U.S. states will use a risk assessment tool to help determine the length of the defendant's sentence. Some of the more sophisticated risk assessments in use today are proprietary computer algorithms created by private companies. These algorithms use statistical probabilities based on factors such as age, employment history, and prior criminal record to predict a defendant's likelihood of recidivism.

In 2016, the investigative journalism non-profit, ProPublica, alleged that a popular risk assessment algorithm called COMPAS was racially biased against black defendants. The corporation behind COMPAS refuted these allegations, arguing that their algorithm predicted recidivism accurately regardless of race. Strangely, both claims proved true. Though COMPAS was fair in one respect, the algorithm was discriminatory under a different definition of fairness. And it was impossible to make the algorithm fair in both ways.

This Note examines the problems underlying risk assessment algorithms used in sentencing and analyzes two key definitions of fairness that can be used in creating and evaluating these algorithms. While the analysis focuses heavily on the COMPAS algorithm, the fairness paradigms and issues of racial discrimination apply to risk assessment algorithms generally.

---

\* J.D. Candidate 2020, Columbia Law School; LL.M. Candidate 2020, University of Amsterdam Law School; B.A. 2016, Northwestern University. I would like to thank Professor Daniel Richman for his steadfast support and incomparable guidance, Professor Crystal Yang for her exceptional insights and encouragement, and the staff of the *Columbia Human Rights Law Review* for their invaluable editorial assistance.

Part I of this Note provides a condensed history of the use of risk assessments in criminal justice and presents recent issues of racial bias in the application of algorithmic risk assessments at sentencing. Part II explores two different definitions of fairness in the controversy surrounding one of the most widely used sentencing algorithms and demonstrates the inherent inability of these definitions to operate in tandem. This phenomenon showcases the opportunity for a different, race-conscious framework in criminal sentencing algorithms. Finally, Part III proposes possible changes to ameliorate racial bias in risk assessment and offer a new fairness framework for sentencing algorithms based on the definition of equalized odds, emulating the model of affirmative action in higher education.

## TABLE OF CONTENTS

Introduction .....	216
I. The Evolution of Risk Assessment Tools in the American Criminal Justice System .....	218
A. Brief History of Criminal Sentencing and Risk Assessment Tools .....	220
1. What Is a Risk Assessment Tool? .....	220
2. Modern Risk Assessments .....	222
B. How Reliable Are Modern Risk Assessments and What Is Their Utility? .....	223
C. Algorithmic Sentencing .....	225
1. <i>State v. Loomis</i> .....	226
2. Biased or Not? .....	227
II. Defining Fairness in Risk Assessment Algorithms .....	228
A. Defining Fairness .....	229
1. Error Rate Balance/Equalized Odds .....	230
2. Accuracy Equity and Predictive Parity .....	231
3. Base Rates of Recidivism .....	232
B. An Inability to Compromise .....	234
C. The Benefits of Equalized Odds .....	235
D. Accounting for Race .....	236
E. Previously Suggested Solutions .....	237
III. Accounting For Race: Equalized Odds-Based Risk Assessments .....	239
A. Adding Race .....	240
B. Weighting Low Risk Scores Greater than High Risk Scores .....	241
C. Ensuring Validation of Algorithms for Race .....	242
D. A More Radical Solution: Affirmative Action for Sentencing Algorithms .....	242
Conclusion .....	244

## INTRODUCTION

A judge sits in her chambers, painstakingly weighing a range of factors in deciding upon the length of a defendant's criminal sentence: the applicable sentencing guidelines recommendations, the harm done to the community, the defendant's criminal history, and pieces of personal information that might explain what led the defendant down this road.<sup>1</sup> All the while, she utilizes these data points as proxies to solve larger problems. She considers the time necessary for the rehabilitation of the defendant himself, incapacitation for public safety, and the prospect of deterring future crimes.<sup>2</sup> Assessing these oft-competing societal interests requires difficult calculations.<sup>3</sup> There are hundreds of conceivably relevant data points, each one impacting the others in unexpected but potentially significant ways.<sup>4</sup>

To alleviate some of the burden, and in the hopes of coming to her conclusion with greater accuracy, the judge turns to a risk-

---

1. DON M. GOTTFREDSON, NAT'L INST. OF JUSTICE, *EFFECTS OF JUDGES' SENTENCING DECISIONS ON CRIMINAL CAREERS* 3 (1999) ("Judges most often reported a crime control aim as the main reason they imposed the sentences they did. Rehabilitation and specific deterrence were prominent considerations . . . . Judges typically had more than one purpose.").

2. See AM. BAR ASS'N, CRIMINAL JUSTICE SECTION, STATE POLICY IMPLEMENTATION PROJECT 18 (2010), [http://www.americanbar.org/content/dam/aba/administrative/criminal\\_justice/spip\\_handouts.authcheckdam.pdf](http://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_handouts.authcheckdam.pdf) [<https://perma.cc/6ZF5-SZDW>].

3. This is especially true when weighing interests such as retribution or moral desert, which are not as easily captured by sentencing data. Mandatory-minimum sentences and Three-Strikes laws are often explained in terms of societal retribution, but do not necessarily help explain the weight assigned to retribution by decisionmakers. Moreover, perceptions of how important retribution is in sentencing varies across surveys and interviews with laypeople, law enforcement, and legal professionals. See James Bernard et al., *Perceptions of Rehabilitation and Retribution in the Criminal Justice System: A Comparison of Public Opinion and Previous Literature*, J. FORENSIC SCI. & CRIM. INVESTIGATION, Oct. 3, 2017, at 6–7. But see Alice Ristrop, *Desert, Democracy, and Sentencing Reform*, 96 J. CRIM. L. & CRIMINOLOGY 1293, 1328–29 (2006) (positing that information regarding the imposition of death sentences can shed light on the role of retributivism in that context, because "capital sentencing decisions are largely the products of inquiries into, and assessments of, the moral desert of the individual defendant.").

4. Brandon M. Greenwell et al., *A Simple and Effective Model-Based Variable Importance Measure*, ARXIV 13–14 (May 15, 2018), <https://arxiv.org/pdf/1805.04755.pdf> [<https://perma.cc/33L6-CXHW>] (describing how a measure of variable importance in a model-based approach can quantify the impact that one variable has on another variable in a data set).

assessment algorithm.<sup>5</sup> This algorithm, having run all of the relevant data points through its multi-level, proprietary actuarial model, will tell the judge how likely the defendant is to recidivate: low risk, medium risk, or high risk.<sup>6</sup> This determination, like a critical piece of the puzzle, gifts the judge a recommended sentence.<sup>7</sup> In theory, this recommended sentence removes individual human bias while maintaining fairness and achieving the broad objectives of justice throughout the system.<sup>8</sup>

This, of course, is an overly idealistic narrative—one that proponents of risk assessment algorithms push when discussing judicial sentencing—giving a shine to this latest manifestation of

---

5. William M. Grove et al., *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 12 PSYCHOL. ASSESSMENT 19, 19 (2000).

6. The legal definition of “recidivate” covers all actions that will result in the defendant returning to prison. This ranges from the defendant committing a new crime to simply breaking the rules of parole or probation. Some recidivism measurements are limited to relatively short timeframes, often three to five years. For further information on measuring recidivism, see *Recidivism*, NAT'L INST. OF JUST. (June 17, 2014), <https://www.nij.gov/topics/corrections/recidivism/Pages/welcome.aspx> [https://perma.cc/C9GW-389X].

7. For an example of how risk assessments “recommend” a sentence, see MICHAEL BAGLIVIO & MARK RUSSELL, THE FLORIDA DEPARTMENT OF JUVENILE JUSTICE DISPOSITION MATRIX: A VALIDATION STUDY 6 (2014), <http://www.djj.state.fl.us/docs/research2/the-fdj-disposition-matrix-validation-study.pdf?sfvrsn=0> [https://perma.cc/RF6E-L26D] (showing the matrix used for sentencing juvenile offenders to prison alternatives).

8. While judges do not rely entirely on an algorithm when determining a sentence, recent research suggests humans may unconsciously trust algorithms more than we realize. One recent survey found that half of Virginia’s state judges rely equally upon the state’s Nonviolent Risk Assessment tool and their judicial experience when making a sentencing decision. Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CALIF. L. REV. (forthcoming 2020). If accurate, assurances that sentencing algorithms play a small role—as simply one of many factors in consideration—become less credible. To be sure, such a finding goes against the traditional scholarship suggesting that humans distrust algorithmic output. See, e.g., Dietvorst et al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCHOL.: GEN., 2015, at 1 (showing that people prefer human forecasters over algorithmic forecasters and that “people more quickly lose confidence in algorithmic than human forecasters after seeing them make the same mistake”) [hereinafter *Algorithmic Aversion*]. And yet, six recent studies found that “participants relied *more* on identical advice when they thought it came from an algorithm than when they thought it came from other people.” Jennifer M. Logg et al., *Algorithm Appreciation: People Prefer Algorithmic to Human Judgment* 14 (Harv. Bus. Sch. Working Paper 17-086, 2018). While “experienced professionals . . . relied less on algorithmic advice than lay people did,” mounting evidence suggests that algorithms can play a bigger role in human decision-making than we have been willing to admit. *Id.* at 2.

evidence-based justice. But that shine can be scrubbed away rather easily. For instance, would you be troubled to know that a popular risk assessment algorithm was more likely to misclassify black defendants as high risk and white defendants as low risk?<sup>9</sup> Or that the public cannot gain access to the methodology by which many risk assessments are built and tested?<sup>10</sup> If so, these revelations might lead to a bigger question: Do sentencing algorithms actually solve the problems they purport to address, or do they simply expose even deeper issues in American society?<sup>11</sup>

Part I of this Note provides a condensed history of the use of risk assessments in criminal justice and presents recent issues of racial bias in the application of algorithmic risk assessments at sentencing. Part II explores two different definitions of fairness in the controversy surrounding one of the most widely used sentencing algorithms and demonstrates the inherent inability of these definitions to operate in tandem. This phenomenon showcases the opportunity for a different, race-conscious framework in criminal sentencing algorithms. Finally, Part III proposes possible changes to ameliorate racial bias in risk assessment and offer a new fairness framework for sentencing algorithms based on the definition of equalized odds, emulating the model of affirmative action in higher education.

## I. THE EVOLUTION OF RISK ASSESSMENT TOOLS IN THE AMERICAN CRIMINAL JUSTICE SYSTEM

Actuarial tools have been used in the realm of criminal justice for nearly a century, but the leap into the digital age and the accompanying increase in computing power has rapidly changed the

---

9. 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> [https://perma.cc/CM8P-Z2DS].

10. Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1022, 1024 (2017) (explaining how corporations protect their algorithms from public scrutiny by employing trade secret law).

11. “Big data helps answer what, not why, and often that’s good enough.” Kenneth Neil Cukier & Viktor Mayer-Schoenberger, *The Rise of Big Data*, FOREIGN AFF., May–June 2013, at 28, 29. In this context, big data and algorithms show “what” patterns exist in a data set and can thereby predict the likelihood of those patterns repeating. But when it comes to understanding the consequences of risk assessments at sentencing, observing patterns is not enough.

landscape of their application.<sup>12</sup> From parole decisions to criminal sentencing, complex algorithms are now being used as tools to make some of the most important decisions in our criminal justice system.<sup>13</sup> Specifically, using algorithmic risk assessment tools at sentencing has become a widespread practice that shows no signs of retreat.<sup>14</sup> And yet, concerns about sentencing uniformity and equality that such algorithms were supposed to put to rest stubbornly remain.<sup>15</sup> In fact, algorithmic risk assessment tools now present us with even more pernicious questions about the workings of the American criminal justice system.<sup>16</sup> For example, what do we mean when we say criminal sentencing should be fair? To whom?

---

12. With the ability to analyze immense quantities of data, risk assessments can be tailored to specific populations—down to individual counties in a state, as is the case with the use of the COMPAS risk assessment in Broward County, Florida. See *infra* p. 12 and note 63. It also allows for risk assessments to analyze more information pertaining to each individual. Cf. *infra* note 23; NORTHPOINTE INC., SAMPLE COMPAS RISK ASSESSMENT 5 (2011), <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html> [https://perma.cc/66FJ-5HF9] (showing that an early risk assessment consisted of 21 data points about a defendant, while Northpointe's COMPAS program is comprised of 137 data points). But see BRIAN J. OSTRUM ET AL., NATIONAL CENTER FOR STATE COURTS, OFFENDER RISK ASSESSMENT IN VIRGINIA 27 (2002), [http://www.vcsc.virginia.gov/risk\\_off\\_rpt.pdf](http://www.vcsc.virginia.gov/risk_off_rpt.pdf) [https://perma.cc/24AS-EGE9]. The Offender Risk Assessment used in Virginia has only 11 factors in evaluating many crimes and is not significantly less accurate than its peers that use more factors.

13. See Judge Irving R. Kaufman, *Sentencing: The Judge's Problem*, ATLANTIC (Jan. 1960), <https://www.theatlantic.com/past/docs/unbound/flashbk/death/kaufman.htm> [https://perma.cc/CKB2-K4VH] (“In no other judicial function is the judge more alone; no other act of his carries greater potentialities for good or evil than the determination of how society will treat its transgressors.”).

14. See Pari McGarraugh, Note, *Up or Out: Why “Sufficiently Reliable” Statistical Risk Assessment Is Appropriate at Sentencing and Inappropriate at Parole*, 97 MINN. L. REV. 1079, 1080 (2013); see also *Algorithms in the Criminal Justice System*, ELEC. PRIVACY INFO. CTR., <https://epic.org/algorithmic-transparency/crim-justice/> [https://perma.cc/L85A-LT8K] [hereinafter EPIC Algorithm List] (providing a summary and background information on risk assessment tools and including a list of such tools used by each state).

15. See Eric Holder, U.S. Attorney General, Speech at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference (Aug. 1, 2014).

16. See Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It’s Actually Not That Clear.*, WASH. POST (Oct. 17, 2016), <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> (on file with the *Columbia Human Rights Law Review*) (asking what it means for an algorithm to be fair and concluding that there is a mathematical limit to fairness).

Section I.A briefly outlines the evolution of actuarial tools in the American criminal justice system. Section I.B explores the accuracy and utility of these tools, uncovering some of the deeply rooted issues with the evidence-based sentencing approach. Section I.C then lays out the current state of play in algorithmic sentencing and introduces the major recent controversy in the field: *State v. Loomis* and the COMPAS risk assessment tool.<sup>17</sup>

#### A. Brief History of Criminal Sentencing and Risk Assessment Tools

##### 1. What Is a Risk Assessment Tool?

An algorithm, at its most basic, is a step-by-step process for solving a problem. Often portrayed as a formula predicting or leading to a future outcome,<sup>18</sup> algorithms range from the exceedingly simple to the startlingly complex.<sup>19</sup> In this case, a risk assessment algorithm is a model that uses “statistical probabilities based on factors such as age, employment history, and prior criminal record” to predict a defendant’s likelihood of recidivism.<sup>20</sup> The hypothetical played out in the Introduction above describes a standard process of how such a risk assessment is presently used in sentencing a criminal defendant. But it was not always so.

As recently as the mid-twentieth century, the law afforded trial judges ample latitude in determining the sentence for a convicted

---

17. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

18. This definition paraphrases a dictionary definition of algorithm. See *Algorithm*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/algorithm> [https://perma.cc/GPC6-MCEE]. Legal scholarship on algorithms sometimes employs more technical language. For example, one paper defined “algorithm” as “any evidence-based forecasting formula or rule. Thus, the term includes statistical models, decision rules, and all other mechanical procedures that can be used for forecasting.” Dietvorst et al., *supra* note 8, at 1.

19. A recipe for baking a cake, for example, is technically a simple algorithm. The several million lines of computer code that make up a quantitative hedge fund’s profit strategy, on the other hand, illustrate the other extreme of algorithmic complexity. See Katherine Burton, *Inside the Medallion Fund, a \$74 Billion Money-Making Machine Like No Other*, FIN. REV. (Nov. 22, 2016), <https://www.afr.com/technology/inside-the-medallion-fund-a-74-billion-moneymaking-machine-like-no-other-20161122-gsuohh> [https://perma.cc/8RKX-2FWP].

20. Anna Maria Barry-Jester et al., *Should Prison Sentences Be Based on Crimes that Haven’t Been Committed Yet?*, FIVETHIRTYEIGHT (Aug. 4, 2015) <https://fivethirtyeight.com/features/prison-reform-risk-assessment/> [https://perma.cc/RC6W-RSNN].

defendant.<sup>21</sup> There was little need to explain how dangerous a defendant was perceived to be, much less to justify the length of a sentence by comparing it to the sentences of other defendants who had committed the same crime. What's more, these sentences were largely unreviewable.<sup>22</sup> Defendants convicted of the same crime and bearing largely similar personal characteristics could thereby receive wildly varying sentencing lengths for no discernible reason.

The earliest risk assessments tools, had they been used at all in mid-twentieth century sentencing, would thus have had little appeal to the freewheeling judges of the time. This is also partly due to the fact that the first iterations were rudimentary at best. In 1927, one of the first risk assessments used twenty-one factors to predict success of parole.<sup>23</sup> This, and subsequent assessments, functioned like checklists. Each factor applied to a defendant would add one point to their score. Higher scores indicated a higher risk of recidivism. While primitive and simple to administer, these basic assessments struggled not only to achieve the desired predictive accuracy, but also to measure and improve upon that accuracy.<sup>24</sup> Throughout the following decades, risk assessments for use at pre-trial and parole hearings proliferated despite continuing to be plagued by inaccuracy. According to one study, early risk assessments "identifying subjects as dangerous were wrong twice as often as they were right."<sup>25</sup> Another study "produced false positive rates of well over 50 percent," incorrectly predicting that offenders would recidivate more often than correctly predicting so.<sup>26</sup>

---

21. See Douglas A. Berman, *Beyond Blakely and Booker: Pondering Modern Sentencing Process*, 95 J. L. & CRIMINOLOGY 653, 654 (2005).

22. Marvin E. Frankel, *Lawlessness in Sentencing*, 41 U. CIN. L. REV. 1, 9 (1972).

23. BERNARD E. HAROURT, *AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE* 58 (2007).

24. *Id.* at 59. For further information on recidivism, see generally *Recidivism*, *supra* note 6 (providing an overview on the definition of recidivism and its role in considering core criminal justice topics); see also Pamela M. Casey et al., *Using Offender Risk and Needs Assessment Information at Sentencing*, NAT'L CTR. FOR STATE COURTS (2011), <https://www.ncsc.org/~media/Microsites/Files/CSI/RNA%20Guide%20Final> [<https://perma.cc/424V-F7XP>] (guiding judges and others involved in sentencing decisions on the appropriate usage of risk and needs assessment instruments).

25. J. C. Oleson et al., *Training to See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessments Among Federal Probation Officers*, 75 FED. PROBATION 52, 52 (2011).

26. Christopher Slobogin, *Risk Assessment and Risk Management in Juvenile Justice*, 27 CRIM. JUST. 10, 12 (2013) (citing JOHN MONAHAN, *PREDICTING VIOLENT BEHAVIOR: AN ASSESSMENT OF CLINICAL TECHNIQUES* 44–49 (1981)).

These persistently high error rates meant that even with the flexible structure of indeterminate sentencing, mid-century risk assessments were not yet ready for use at sentencing.

## 2. Modern Risk Assessments

The second half of the twentieth century brought an end to indeterminate sentencing and briefly ushered in the era of mandatory sentencing guidelines, further reducing the usefulness of risk assessments.<sup>27</sup> But as state and federal guidelines were loosened at the turn of the millennium, judicial discretion saw a resurgence. To handle this, judges at both the state and federal levels increasingly turned to a host of new tools for analyzing the information relevant for sentencing.<sup>28</sup> Though operating under different names and measuring slightly different categories,<sup>29</sup> these tools retain the essence of the original risk assessments. Moreover, they still all appear to have at least one goal in common with their earlier iterations: predicting the probability that a defendant will recidivate.<sup>30</sup>

---

27. *But see* Kevin R. Reitz, *The Enforceability of Sentencing Guidelines*, 58 STAN. L. REV. 155, 155–56 (2006) (placing the federal system and 18 state guidelines systems on a continuum ranging from advisory guidelines with no requirement for a statement justifying departure from the guidelines to the “mandatory” pre-*Booker* federal guidelines).

28. *See EPIC Algorithm List*, *supra* note 14. The list provides detailing information on risk assessment tools used state-by-state and whether a state has conducted a validity test on the tool in use. While risk assessments like COMPAS are used both in the pre-trial setting and at sentencing, this Note will focus exclusively on the use of risk assessments at the sentencing phase. *See, e.g.*, *Our Products*, EQUIVANT, <https://www.equivant.com/classification/> [<https://perma.cc/B2T2-YR9S>] (examples of various criminal justice risk assessment tools).

29. *EPIC Algorithm List*, *supra* note 14, at Background. COMPAS assesses variables under five main areas: criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion. The LSI-R, another risk assessment tool, also pulls information “ranging from criminal history to personality patterns.” The Public Safety Assessment, however, “only considers variables that relate to a defendant’s age and criminal history.” *Id.*

30. For more information on the evolution of risk assessment tools, often categorized into “generations,” *see* Danielle Kehl et al., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, RESPONSIVE COMMUNITIES, HARV. L. SCH. (2017), <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041> [<https://perma.cc/XBJ7-QG47>]; *see also* Recent Cases, *Criminal Law—Sentencing Guidelines—Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*, 130 HARV. L. REV. 1530, 1530 (2017) (analyzing Wisconsin Supreme Court opinion holding that not disclosing mechanism of risk assessment to defendant did not violate defendant’s due process rights, discussed further *infra* Section I.C.1).

To achieve this goal, creators of these tools gather data on large populations of former prisoners and then track these prisoners for years “to see which traits are associated with further criminal activity.”<sup>31</sup> Going forward, these tools determine the relative predictive weight of such traits and apply them to the individual in question, reaching a rough “score” of how likely that individual is to reoffend.<sup>32</sup> Proponents of these risk assessments argue that such an approach more accurately predicts recidivism than an individual judge, while lowering budget costs and saving time for all actors in the criminal justice system.<sup>33</sup>

### B. How Reliable Are Modern Risk Assessments and What Is Their Utility?

Recently, however, the accuracy of risk assessment tools has again been called into question.<sup>34</sup> According to Professor Sonja Starr at Michigan Law School, “there is no persuasive evidence that the instruments’ predictive power exceeds that of either the current system . . . or less discriminatory alternative instruments.”<sup>35</sup> Corroborating this claim, recent studies of “presumptively high-risk populations” still produce false positive rates ranging from 15% to 50%.<sup>36</sup> While this marks an improvement from decades ago, such persistently high rates are troubling, suggesting that risk assessment tools are still no better than the humans they intend to replace.<sup>37</sup> A

---

31. Barry-Jester et al., *supra* note 20.

32. *Id.*

33. ANGÉLE CHRISTIN ET AL., COURTS AND PREDICTIVE ALGORITHMS 1–2 (Oct. 27, 2015), [https://datasociety.net/wp-content/uploads/2015/10/Courts\\_and\\_Predictive\\_Algorithms.pdf](https://datasociety.net/wp-content/uploads/2015/10/Courts_and_Predictive_Algorithms.pdf) [<https://perma.cc/J73B-A5MA>].

34. Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 806–07 (2014).

35. *Id.* at 807.

36. Slobgin, *supra* note 26, at 11. In this scenario, a false positive rate is the fraction of defendants incorrectly identified as individuals at high risk for recidivating.

37. It is worth remembering that a risk assessment tool’s recommended sentence, which rests entirely on predictions of recidivism, is qualitatively different from other determinations leading to a judge’s final sentence. Judges must consider many factors when fashioning a sentence, making recidivism only one element of their calculations. See *supra* notes 1–4 and accompanying text. Consequently, a judge’s ability to accurately predict recidivism has less of an impact on her final sentence than does a risk assessment tool’s accuracy on its recommended sentence. Comparing the accuracy of a judge and a risk assessment tool is thus helpful for understanding how capable a risk assessment may be at replacing human

recent study by Julia Dressel and Hany Farid purports to show that the COMPAS<sup>38</sup> risk assessment in Wisconsin “is no more accurate or fair” than the predictions of individual laypeople on the Internet.<sup>39</sup> Even more alarming, simply pooling the responses of the study’s participants returned a greater accuracy rate than that of the COMPAS algorithm.<sup>40</sup>

As this issue has itself shifted further into the spotlight, questions of accuracy have been accompanied, if not replaced, by questions about bias. While factors like socioeconomic status, employment status, and marital status may have predictive value, their usage in risk assessments also “explicitly endors[es] sentencing discrimination based on factors the defendant cannot control.”<sup>41</sup> For example, information about where a defendant lives can “penalize residents of urban areas, who are far more likely to be black.”<sup>42</sup> Furthermore, factoring a defendant’s zip code into the assessment may indirectly account for racially biased policing practices in their neighborhood, thereby compounding institutionalized discrimination from earlier in the criminal justice process.<sup>43</sup>

Even subjective measures on a risk assessment can count against a defendant. In *State v. Gauthier*, the Supreme Judicial Court of Maine found Gauthier’s high score on a popular risk assessment tool

---

predictions of recidivism, but even this comparison does not go so far as implicating whether a risk assessment tool is capable of replacing a judge at sentencing.

38. COMPAS is an abbreviation for Correctional Offender Management Profiling for Alternative Sanctions. *On Demand Trainings*, GLOBAL INST. OF FORENSIC RES., <https://www.gifrinc.com/course/compas/> [<https://perma.cc/F436-8ULJ>].

39. Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES 1–2 (Jan. 17, 2018), <https://advances.sciencemag.org/content/4/1/eaao5580/tab-pdf> [<https://perma.cc/4GZ7-9XJN>]. Each participant was randomly assigned 50 defendants and tasked with predicting recidivism within two years. The “[p]articipants saw a short description of a defendant that included the defendant’s sex, age, and previous criminal history, but not their race.” The average of the median participant accuracy was 62.8%, less than one standard deviation away from the accuracy rate of the COMPAS risk assessment at 65.2%. *Id.*

40. Ed Yong, *A Popular Algorithm Is No Better at Predicting Crimes than Random People*, ATLANTIC (Jan. 17, 2018), <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/> [<https://perma.cc/4HL9-WFE9>].

41. Sonja B. Starr, *The New Profiling: Why Punishing Based on Poverty and Identity Is Unconstitutional and Wrong*, 27 FED. SENT’G REP. 229, 230 (2015).

42. Barry-Jester et al., *supra* note 20.

43. Jeffrey Fagan & Daniel Richman, *Understanding Recent Spikes and Longer Trends in American Murder Rates*, 117 COLUM. L. REV. 1235, 1247 (2017).

to be an aggravating factor in his sentencing.<sup>44</sup> Despite Gauthier's youth and history of mental illness, the Court specifically noted his "lack of respect for others" and "refusal to accept responsibility."<sup>45</sup> Combined with the "Emotional/Personal" and "Attitude/Orientations" categories on the risk assessment, such subjective determinations contributed to Gauthier's sixty-year sentence for murder.<sup>46</sup> In this light, glaring problems of accuracy, bias, and subjectivity remain unsolved. Even the most modern and sophisticated risk assessment algorithms today are plagued by issues of discrimination.

### C. Algorithmic Sentencing

Increasingly, modern risk assessment algorithms rely on what has come to be called "big data." A rather flexible term, big data generally describes a large volume of data that can be mined for information. Crucially, the term big data implies that by analyzing a massive amount of data correctly, we can learn much more than we previously could with less data.<sup>47</sup> As one might imagine, the main distinction between algorithms today and risk assessment tools of the past is their use of big data.<sup>48</sup>

By gathering decades worth of publicly available information at the individual, community, and state levels, these algorithms are able to make predictions about future crime in an entirely different manner than risk assessment tools that simply add up points from a checklist.<sup>49</sup> Another crucial element here is that these algorithms are *automated*. As Professor Aziz Huq at the University of Chicago Law School has defined it, algorithmic criminal justice is "the application of an automated protocol to a large volume of data" in order to, amongst

---

44. State v. Gauthier, 989 A.2d 77, 85 (Me. 2007).

45. *Id.* at 81, 86.

46. *Id.*

47. Cukier & Mayer-Schoenberger, *supra* note 11, at 28–30.

48. "As recently as the year 2000, only one-quarter of all the world's stored information was digital. . . . Today, less than two percent of all stored information is nondigital." *Id.* at 28–29; see also *Forecast of Big Data Market Size, Based on Revenue, from 2011 to 2027 (in Billion U.S. Dollars)*, STATISTA (Aug. 9, 2019), <https://www.statista.com/statistics/254266/global-big-data-market-forecast/> [<https://perma.cc/E643-SGJF>] (showing growth of big data from 2011–2017 and projected growth from 2018–2027).

49. Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947, 966 (2016) (citing Christopher Slobogin, *Risk Assessment*, in THE OXFORD HANDBOOK OF SENTENCING AND CORRECTIONS 196, 200 (Joan Petersilia & Keven R. Reitz eds., 2012)).

other things, “make out-of-sample predictions about new actors’ likely criminal conduct.”<sup>50</sup>

While automation is necessary to process massive quantities of data,<sup>51</sup> this also means that a recommended criminal sentence can be calculated in a black box.<sup>52</sup> Naturally, such a dearth of information about risk assessment algorithms’ inner workings raises questions from individuals, courts, and the public.<sup>53</sup> Foremost among these questions might be: “Why are we allowing a computer program, into which no one in the criminal justice system has any insight, to play a role in sending a man to prison?”<sup>54</sup>

### 1. *State v. Loomis*

This question was recently brought to life in the Wisconsin case of *State v. Loomis*.<sup>55</sup> In 2013, Eric Loomis, charged with five criminal counts as the driver in a drive-by shooting, was identified as a high risk individual by the previously mentioned COMPAS risk assessment used in Wisconsin.<sup>56</sup> On appeal, he filed a motion for post-conviction relief on the grounds that use of COMPAS at sentencing violated his due process rights.<sup>57</sup> As an expert witness testified, “[t]he Court does not know how the COMPAS compares that individual’s history with the population that it’s comparing them with. The Court doesn’t even know

---

50. Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1060–61 (2019).

51. *Id.* at 1061.

52. Ellora Thadaney Israni, Opinion, *When An Algorithm Helps Send You to Prison*, N.Y. TIMES (Oct. 26, 2017), <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html> (on file with the *Columbia Human Rights Law Review*) (“No one knows exactly how COMPAS works; its manufacturer refuses to disclose the proprietary algorithm. We only know the final risk assessment score it spits out, which judges may consider at sentencing.”).

53. Barry-Jester et al., *supra* note 20 (including interviews with a man who served jail time for a DUI who says the use of predictive algorithms in sentencing “ain’t right” and a Michigan law school professor who says it is not “fair,” and citing evidence that “judges disregard sentencing guidelines roughly 20 percent of the time”); *see also* State v. Loomis, 881 N.W.2d 749, 757 (Wis. 2016) (explaining defendant Loomis’ due process challenge against the use of COMPAS at sentencing, which was based in part on the fact that he could not gain insight into how the recommended sentence was formulated due to the proprietary nature of the algorithm).

54. Israni, *supra* note 52.

55. 881 N.W.2d 749 (Wis. 2016).

56. *Id.* at 754–55.

57. *Id.* at 756.

whether that population is a Wisconsin population, a New York population, a California population . . . .”<sup>58</sup>

The court could not obtain such information due to the proprietary nature of the COMPAS algorithm and the trade secret protections acquired by its developer, Northpointe, Inc.<sup>59</sup> Nevertheless, the Wisconsin Supreme Court upheld the use of COMPAS, subject to certain limitations.<sup>60</sup> Wisconsin circuit courts would be required to explain the other factors involved in their sentencing decisions, and any COMPAS report would need to be accompanied by a disclaimer regarding the accuracy and appropriate uses of COMPAS.<sup>61</sup>

## 2. Biased or Not?

In the months between the time *State v. Loomis* was argued and decided,<sup>62</sup> the investigative journalism non-profit, ProPublica, published a scathing article about COMPAS, alleging that the algorithm was racially biased against black defendants.<sup>63</sup> This conclusion stemmed from analyzing a data set of 10,000 prisoners in Broward County, Florida, for which COMPAS had created risk scores.<sup>64</sup> In short, ProPublica found that the algorithm incorrectly predicted black defendants to be high risk more often than white defendants, while also incorrectly predicting white defendants to be lower risk more often than black defendants.<sup>65</sup> While “the algorithm correctly predicted recidivism for black and white defendants at

---

58. *Id.*

59. *Id.* at 760.

60. *Id.* at 763. The court held that Loomis’ ability to review the risk score itself, if not the underlying calculation of the score, sufficed to uphold his “due process right to be sentenced based on accurate information.” *Id.* at 760–61. To more effectively safeguard this right, however, Judge Bradley also held that any sentencing court in Wisconsin must “explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed.” *Id.* at 769. Finally, courts must include five disclaimers with any Presentence Investigation (PSI) report containing a COMPAS risk assessment to emphasize the limitations of the risk assessment. *Id.*

61. *Id.*

62. The case was argued on April 5, 2016. *Id.* at 749. ProPublica published its report on May 23, 2016. See *infra* note 63. The Wisconsin Supreme Court then issued its ruling on July 13, 2016. *Loomis*, 881 N.W.2d at 749.

63. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<https://perma.cc/2G2Y-DJRX>].

64. Larson et al., *supra* note 9.

65. *Id.*

roughly the same rate,” its mistakes manifested in opposite ways with respect to race.<sup>66</sup>

Northpointe quickly responded, refuting ProPublica’s findings of racial bias in the COMPAS algorithm.<sup>67</sup> Northpointe’s main assertion was that ProPublica’s analysis “did not take into account the different base rates of recidivism for blacks and whites.”<sup>68</sup> ProPublica soon issued their own rebuttal to Northpointe’s allegations, defending their logistic regression and standing by their findings.<sup>69</sup> However, what may have been lost amid the intricacies of this statistical debate is that, in a way, both ProPublica and Northpointe can be correct. And they can both be wrong. It all depends on how you define fairness.<sup>70</sup>

## II. DEFINING FAIRNESS IN RISK ASSESSMENT ALGORITHMS

Section II.A of this Part explores the relevant definitions of fairness in the context of Northpointe and ProPublica’s dispute. Section II.B then reveals the incompatibility of these definitions of fairness given the data on race and crime in America. Section II.C highlights the benefits of using the equalized odds definition of fairness in this context, while Section II.D investigates the current use of race in algorithms and how proxies are used. Finally, Section II.E surveys previously suggested solutions for reducing racial bias in the algorithms, ultimately showing that they have not succeeded. In this light, the Note demonstrates that there is space for a new solution based on equalized odds and race-consciousness.

---

66. *Id.* at 1–2. The analysis revealed a 59% correct prediction rate for white defendants and a 63% correct prediction rate for black defendants. However, it also found high risk misclassifications at a rate of 45% for black defendants and 23% for white defendants, and low risk misclassifications at 48% for white defendants and 28% for black defendants. In other words, black defendants were misclassified as high risk at twice the rate that white defendants were misclassified as high risk, while white defendants were misclassified as low risk at nearly twice the rate that black defendants were misclassified as low risk. *Id.*

67. WILLIAM DIETERICH ET AL., NORTHPOINTE, INC. RESEARCH DEPT’, COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY 2 (2016).

68. *Id.* at 1.

69. Jeff Larson & Julia Angwin, *Technical Response to Northpointe*, PROPUBLICA (July 29, 2016), <https://www.propublica.org/article/technical-response-to-northpointe> [<https://perma.cc/7UZZ-CW4B>].

70. Corbett-Davies et al., *supra* note 16, at 2.

### A. Defining Fairness

Fairness is a difficult concept on which to pin precise definitions and theories. Nonetheless, clear and distinct definitions of fairness are critical to understanding and potentially correcting the injustices that are currently a byproduct of risk assessment algorithms. Computer scientists of all stripes have been learning this lesson the hard way, struggling to generate and apply definitions of fairness in increasingly complex algorithms for fields such as artificial intelligence and autonomous vehicles.<sup>71</sup> Yet, the academic computer science community has made serious strides in recent years towards devising narrow and actionable definitions of fairness that are now being applied in social science and legal scholarship.<sup>72</sup> While there are numerous definitions, each undergirded by rigorous peer-reviewed research,<sup>73</sup> two definitions are most salient for the purposes of this

---

71. Jeanette Wing, a pioneer in the field of fairness and ethics in data science and one of the first leading data scientists to identify the ethical dilemmas in data-driven decision-making, has developed an acronym that provides a framework for addressing these issues: FATES (Fairness, Accountability, Transparency, Ethics, Safety and Security). Jeanette Wing, *Data for Good*, DATA SCI. INST. (Jan. 23, 2018), <https://datascience.columbia.edu/data-for-good> [https://perma.cc/4YTY-7ASR]. This framework can be used to address major dilemmas facing modern algorithms, such as racial bias in algorithms or the classic “trolley problem” as applied to self-driving cars. Arian Marshall, *What Can the Trolley Problem Teach Self-Driving Car Engineers?*, WIRED (Oct. 24, 2018), <https://www.wired.com/story/trolley-problem-teach-self-driving-car-engineers/> [https://perma.cc/5KSA-EWJG].

72. Clarifications of the various notions of “fairness” in the law have recently come from researchers in the computer science community. See generally Cynthia Dwork et al., *Fairness Through Awareness*, ARXIV (2011), <https://arxiv.org/pdf/1104.3913.pdf> [https://perma.cc/S6KJ-3DVP] (investigating the mathematical differences between individual fairness and group fairness and illustrating how a program or initiative understood as “fair” to a group of people may be “unfair” to certain individuals, and vice-versa). As Solon Barocas and others have shown, one way “fairness” can be conceptualized is by “inducing a rule from an entire population’s behavior” and attempting to apply that rule to specific individuals. Solon Barocas et al., *Governing Algorithms: A Provocation Piece* 6–7 (N.Y.U. Governing Algorithms Conference, Mar. 29, 2013). Dwork, Barocas, and other computer scientists thereby problematized common-sense notions of fairness, exposing bias arising from schemes with debatable definitions of fairness. To solve problems such as those inherent in COMPAS, social scientists and legal scholars thus have had to more precisely define what “fairness” means in a given context.

73. See, e.g., Alexandra Chouldechova, *Fair Prediction with Disparate Impact*, 5 BIG DATA 153, 154–55 (2017) (defining and comparing “calibration,” “predictive parity,” “error rate balance,” and “statistical parity”); Sam Corbett-Davies & Sharad Goel, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, ARXIV 1–2 (Aug. 14, 2018), <https://arxiv.org/>

subject: equalized odds and predictive parity. Each is highly intuitive, yet distinct enough that Northpointe and ProPublica each claimed one in their quests to demonstrate their adherence to fairness and their adversary's inability to avoid bias.

### 1. Error Rate Balance/Equalized Odds

ProPublica initially alleged that the COMPAS algorithm made classification errors unequally, to the detriment of black defendants.<sup>74</sup> This is known as failure of error rate balance, or failure to equalize odds. In this scenario, for the definition of equalized odds to be satisfied, black and white defendants must have the same odds of being misclassified as High Risk, and the same odds of being misclassified as

---

pdf/1808.00023.pdf [<https://perma.cc/5VDW-JXQT>] (defining and comparing “anti-classification,” “classification parity,” and “calibration”); Sahil Verma & Julia Rubin, *Fairness Definitions Explained* (2018 ACM/IEEE Int’l Workshop on Software Fairness, May 29, 2018) (defining, categorizing, and classifying 20 definitions of fairness); Jon Kleinberg et al., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, ARXIV 3 (2016), <https://arxiv.org/pdf/1609.05807.pdf> [<https://perma.cc/B4ND-BUVC>] (defining and comparing three conditions for fairness: “calibration within groups,” “balance for the negative class,” and “balance for the positive class”).

74. Larson et al., *supra* note 9 (“Our analysis of . . . COMPAS . . . found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk.”). Northpointe takes issue with ProPublica basing certain assertions of bias on interpretations of the data where the cutoff was Low Risk. Julia Angwin & Jeff Larson, *Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*, PROPUBLICA (Dec. 30, 2019), <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say> [<https://perma.cc/Q5E7-UVPJ>]. By regrouping defendants into Low Risk and Higher Risk, ProPublica obtained results that showed greater racial disparity than if the cut off was at Medium Risk, or if there was no cut off and all three original risk classifications were maintained. This appears to be a variation of Simpson’s Paradox (or the Yule-Simpson effect), wherein “the marginal association between two categorical variables is qualitatively different from the partial association between the same two variables after controlling for one or more other variables.” Bruce W. Carlson, *Simpson’s Paradox*, BRITANNICA (Jan. 7, 2019), <https://www.britannica.com/topic/Simpsons-paradox> [<https://perma.cc/5PD7-TE4L>]. While relevant to framing the debate about racial bias in the COMPAS algorithm, the Paradox does not ultimately alter the conclusions surrounding definitions of fairness in this debate. For another relevant example of Simpson’s paradox, see P.J. Bickel et al., *Sex Bias in Graduate Admissions: Data from Berkeley*, 187 SCIENCE 398 (1975). For the pre-eminent theoretical work on this effect, see G. Udny Yule, *Notes on the Theory of Association of Attributes in Statistics*, 2 BIOMETRIKA 16 (1903).

Low Risk. To make this more concrete, it is helpful to examine what group of defendants the definition of equalized odds most benefits.<sup>75</sup>

In ProPublica's analysis of the COMPAS data in Broward County, “[b]lack defendants who do not recidivate were nearly twice as likely to be classified by COMPAS as higher risk compared to their white counterparts (45 percent vs. 23 percent).”<sup>76</sup> That is, in a sample of 7214 defendants, roughly 384 more black defendants than white defendants were misclassified as High Risk.<sup>77</sup> Put another way, the failure of the COMPAS algorithm to equalize odds along racial lines led to 10% of black defendants (384 of the defendants in the sample) being unfairly disadvantaged in their risk assessment score. Thus, up to 10% of black defendants could be benefited—reclassified as Low or Medium Risk—by an algorithm operating under equalized odds.

## 2. Accuracy Equity and Predictive Parity

When Northpointe refuted the allegations of racial bias in the COMPAS algorithm, the company did not dispute these findings of error rate imbalance.<sup>78</sup> Instead, Northpointe asserted that COMPAS was not racially biased because it predicted recidivism and non-recidivism equally well for black and white defendants. Moreover, Northpointe noted that “the probability of recidivating, given a high risk score, is similar for blacks and whites.”<sup>79</sup>

These concepts are defined respectively as “accuracy equity” and “predictive parity.” To satisfy these definitions, an algorithm must base its conception of fairness upon the notion that because it predicts recidivism at the same level of accuracy, regardless of race, its positive

---

75. “Benefits” meaning that though these defendants were misclassified as High Risk under the definition of predictive parity, they would not be misclassified under the definition of equalized odds.

76. Larson et al., *supra* note 9.

77. *Id.* 805 of 1795 High Risk black defendants did not recidivate, while 349 of 1488 High Risk white defendants did not recidivate. Adjusted for equal denominators, 421 of 1795 white defendants would not recidivate.  $805 - 421 = 384$ . Dividing by the total number of black defendants in the sample, 3696, leads to the conclusion that 10% of all black defendants are misclassified as High Risk.

78. Rather, Northpointe argued that “ProPublica focused on classification statistics that did not take into account the different base rates of recidivism for blacks and whites.” DIETERICH ET AL., *supra* note 67, at 1.

79. *Id.* at 35.

predictive value is not racially biased.<sup>80</sup> According to ProPublica's regression analyses, COMPAS exhibited a roughly 62% accuracy rate in predicting recidivism for both black defendants and white defendants, while maintaining a 63.6% accuracy rate across all risk scores.<sup>81</sup> From this angle, Northpointe could validly claim that "[their] test that is correct in equal proportions for all groups cannot be biased."<sup>82</sup>

### 3. Base Rates of Recidivism

Then, in the second half of 2016, computer scientists from multiple prestigious American universities found that given the COMPAS data set, no algorithm could satisfy both definitions of fairness to which ProPublica and Northpointe had subscribed.<sup>83</sup> The problem was simply that black and white defendants had different base rates of recidivism. As the researchers themselves explained:

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.<sup>84</sup>

---

80. Positive predictive value is one measure of accuracy. Generally, it measures the number of times a positive value (here, a prediction that a defendant will recidivate) is the true value (the defendant actually recidivates). For further illustration of this concept, albeit with medical terminology, see *Predictive Value Theory*, U. IOWA, [https://www.healthcare.uiowa.edu/path\\_handbook/appendix/chem/pred\\_value\\_theory.html](https://www.healthcare.uiowa.edu/path_handbook/appendix/chem/pred_value_theory.html) [<https://perma.cc/48XX-M9SC>].

81. Larsen et al., *supra* note 9. Predictive accuracy for recidivism can be measured in a number of ways. When using a Cox regression analysis for the low, medium, and high risk scores, ProPublica found COMPAS had a concordance score of 63.6%. *Id.* Applying the Cox model to the underlying risk scores (using the actual numerical score rather than the low, medium, or high categorization), accuracy increased to 66.4%. *Id.*; see also *supra* note 74 and accompanying text (explaining further how these different classifications can change the predictive accuracy). Finally, in its own study, Northpointe reported a concordance of roughly 68%. Dieterich et al., *supra* note 67, at 3.

82. Angwin & Larson, *supra* note 74.

83. Kleinberg et al., *supra* note 73, at 4.

84. Corbett-Davies et al., *supra* note 16, at 3.

Indeed, analyses of the FBI's Uniform Crime Reporting Statistics confirm differing rates of recidivism along racial lines.<sup>85</sup> These differences are apparent even among juvenile offenders.<sup>86</sup> But if the statistical differences are clear, the root causes of such differences are anything but. Potential causes of differing base rates of recidivism among blacks and whites are far-ranging, though clear front-runners are greater levels of policing in predominantly black communities and higher tendencies for police to make arrests in those neighborhoods.<sup>87</sup>

Looking for explanations outside the realm of law enforcement interactions, the literature on differing base rates of recidivism along racial lines is complex. Even so, a number of studies and accompanying explanations offer a relatively coherent picture of the potential factors creating the divide. One study found that rates of black family disruption "was significantly related to rates of black murder and robbery, particularly by juveniles."<sup>88</sup> Joblessness and poverty were both correlated with family disruption, highlighting why those factors may "have had weak or inconsistent direct effects on violence rates in past research. These factors, in fact, exert influence on family disruption, which in turn, directly affects juvenile violence rates."<sup>89</sup> In

---

85. Robert J. Sampson & Janet L. Lauritsen, *Racial and Ethnic Disparities in Crime and Criminal Justice in the United States*, 21 CRIM. & JUST. 311, 325 (1997) ("For example, in 1993 blacks comprised 31 percent of total arrests yet constituted 12 percent of the population . . .").

86. DARNELL F. HAWKINS ET AL., U.S. DEP'T OF JUSTICE, RACE, ETHNICITY, AND SERIOUS AND VIOLENT JUVENILE OFFENDING 2 (2000), <https://www.ncjrs.gov/pdffiles1/ojjdp/181202.pdf> [<https://perma.cc/4SCS-9NZ6>].

87. "[E]vidence of racially disparate police enforcement across cities reinforces longstanding beliefs among Black citizens about disparate treatment at the hands of the police and helps spread a narrative of an uneven burden that Black citizens bear in police–citizen encounters." Fagan & Richman, *supra* note 43, at 1247. Though Professors Fagan and Richman focus largely on data on the use of force, data on police stops in New York City affirms that even minor police-citizen encounters occur more often in predominantly Black and Latino neighborhoods. See Matthew Bloch et al., *Stop, Question, and Frisk in New York Neighborhoods*, N.Y. TIMES (July 11, 2010), <https://archive.nytimes.com/www.nytimes.com/interactive/2010/07/11/nyregion/20100711-stop-and-frisk.html> (on file with the *Columbia Human Rights Law Review*); see also Dashiell Bennett, 'Stop and Frisk' Continues to Target New York's Poorest People, ATLANTIC (July 3, 2012) <https://www.theatlantic.com/national/archive/2012/07/stop-and-frisk-continues-target-new-yorks-poorest-people/326329/> [<https://perma.cc/T3QU-FXH6>] (citing that, in New York City, police stops are more frequent in the neighborhoods with the poorest and largely minority residents).

88. Sampson & Lauritsen, *supra* note 85, at 335.

89. HAWKINS ET AL., *supra* note 86, at 4.

another study of over 200 urban neighborhoods, researchers found that more highly segregated neighborhoods had higher crime rates.<sup>90</sup>

The interplay of varying potential causes of differing base rates of recidivism is far from conclusive. It does indicate, however, that “discrimination [in the criminal justice system] appears to be indirect, stemming from the amplification of initial disadvantages over time.”<sup>91</sup> Moreover, this lack of clarity serves as a reminder that “an algorithm is only as good as the data it works with.”<sup>92</sup> A tool relying on data infused with a history of racial bias cannot help but reflect that same racial bias back into the society in which it is situated.

### B. An Inability to Compromise

The underlying problem, then, is not that the COMPAS algorithm has been intentionally or consciously coded to discriminate based on race, but that given the state of crime and policing in the United States, a criminal sentencing algorithm cannot boast both predictive accuracy and equalized odds.<sup>93</sup> To be “fair” to some defendants, COMPAS and its ilk must be “unfair” to others. Thus, if it is impossible to be both accurate and equal at the same time, we must choose the definition of fairness by which COMPAS and other risk assessment tools will abide.

---

90. Interview by Resource Center for Minority Data staff with Ruth Peterson and Lauren Krivo (2000), [https://www.icpsr.umich.edu/icpsrweb/content/RCMD/interviews/peterson\\_krivo.html](https://www.icpsr.umich.edu/icpsrweb/content/RCMD/interviews/peterson_krivo.html) [<https://perma.cc/T59N-M3DN>] (discussing the findings of the National Neighborhood Crime Study).

91. Sampson & Lauritsen, *supra* note 85, at 311.

92. Solon Barcas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671, 671 (2016).

93. The author has tried to phrase this claim carefully, because certain readers may take issue with the concept of an algorithm, created by humans and with no agenda of its own, to itself be “biased” in any sense. Despite sensational headlines such as, “*Computer Program that Calculates Prison Sentences Is Even More Racist than Humans*,” the bias in algorithms is either a product of human error or a manifestation of bias already present in human society. See PUB. AFFAIRS, U.C. BERKELEY, *Mortgage Algorithms Perpetuate Racial Bias in Lending, Study Finds*, BERKELEY NEWS (Nov. 13, 2018), [https://news.berkeley.edu/story\\_jump/mortgage-algorithms-perpetuate-racial-bias-in-lending-study-finds/](https://news.berkeley.edu/story_jump/mortgage-algorithms-perpetuate-racial-bias-in-lending-study-finds/) [<https://perma.cc/U6BX-7G88>]. But see Kelly Weill, *Computer Program that Calculates Prison Sentences Is Even More Racist than Humans, Study Finds*, DAILY BEAST (Jan. 21, 2018), <https://www.thedailybeast.com/computer-program-that-calculates-prison-sentences-is-even-more-racist-than-humans-study-finds> [<https://perma.cc/8ZL4-DV8S>] (example of sensational headline).

At present, accuracy equity and predictive parity are the default definitions of fairness under which risk assessment tools can be expected to operate.<sup>94</sup> Of course, there are distinct benefits to operating under such definitions. Given that the main objective of risk assessment tools is to predict recidivism as accurately as possible, prioritizing an algorithm that is well-calibrated is in line with legislative and judicial goals. Accuracy equity allows courts, jails, and prisons to better budget and plan for incarceration and supervision necessities.<sup>95</sup> Finally, predictive parity ensures that risk assessment tools explicitly treat all defendants equally, regardless of race.

### C. The Benefits of Equalized Odds

Nevertheless, the performance of COMPAS in Broward County shows that ensuring equal treatment under accuracy equity and predictive parity still fails to address the ensuing disparate misclassifications. A focus on equalized odds could resolve this issue, leading to an algorithm that operates under a model of disparate treatment to ensure more equal outcomes.

Pursuing equalized odds also aligns with the founding ideals of the American justice system. It is often said that, “it is better that ten guilty persons escape, than that one innocent suffer.”<sup>96</sup> From William Blackstone and Benjamin Franklin to current legal scholars, this principle has long carried weight in considerations of justice. Though COMPAS is but one example, it shows how subtly and forcefully we have shifted from Blackstone’s ideal to its antithesis: allowing innocents to suffer in the hopes that one less crime might be committed.

---

94. There is no publicly available indication that COMPAS has changed its algorithm.

95. See AM. L. INST., MODEL PENAL CODE: SENTENCING, TENTATIVE DRAFT NO. 2, 56 (2011) [hereinafter MPC SENTENCING DRAFT 2 (2011)].

96. 4 WILLIAM BLACKSTONE, COMMENTARIES \*358. Though popularly attributed to Blackstone and referred to as “Blackstone’s ratio,” an earlier version of this sentiment was published by Voltaire: “[T]is much more Prudence to acquit two Persons, tho’ actually guilty, than to pass Sentence of Condemnation on one that is virtuous and innocent.” FRANCOIS-MARIE AROUET VOLTAIRE, ZADIG 53 (1749). This ideal was then carried across the Atlantic by Benjamin Franklin: “That it is better 100 guilty Persons should escape than that one innocent Person should suffer, is a Maxim that has been long and generally approved.” Letter from Benjamin Franklin to Benjamin Vaughan (Mar. 14, 1785), in 9 THE WRITINGS OF BENJAMIN FRANKLIN 293 (Albert H. Smyth ed., 1906).

Certain definitions will be more applicable in attempting to achieve certain goals, and equalized odds may not be the best in each scenario, but the ideals of our justice system and the potential benefits of equalized odds for defendants suggest that equalizing odds is at least as valid as adhering to predictive parity—especially when the unequal odds currently fall along racial lines. Because of this, it is worth examining in greater detail how the law might facilitate and implement risk assessment tools adhering to an equalized odds definition of fairness.

#### D. Accounting for Race

While many of these tools account for the same criminogenic factors and bits of personal data, not one risk assessment tool explicitly incorporates race. As race is a constitutionally protected category, neither federal nor state governments can legally base decisions on race unless doing so would achieve a compelling state interest using the least restrictive, narrowly tailored means.<sup>97</sup> Notwithstanding this, race holds a relatively strong correlation with the risk of recidivism, which, if accounted for, would make it a particularly helpful data point for improving the accuracy of risk assessment tools.<sup>98</sup> Race was actually an explicit factor in the earliest risk assessments aimed at predicting parole violation. The practice of overtly using race to assess risk continued for decades, but for the last half-century it has generally been presumed to be unconstitutional.<sup>99</sup>

Today, instead of explicitly using race as a data point, corporations such as Equivant and MHS Assessments<sup>100</sup> settle for

---

97. *Grutter v. Bollinger*, 539 U.S. 306, 326 (2003).

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

99. *Id.* at 238. Race was used in the earliest risk assessments for predicting parole violation. Recidivism researchers continued to use race as a variable through at least 1967. *See id.* at 241–42 app.

100. MHS Assessments is the owner and proprietor of the LSI-R risk assessment tool. Its website touts it as “the most widely used and widely researched risk/needs assessment in the world.” MHS ASSESSMENTS, LSI-R: LEVEL OF SERVICE INVENTORY-REVISED, [https://www.assessments.com/assessments\\_documentation/LSI-R%20Technical%20Brochure.pdf](https://www.assessments.com/assessments_documentation/LSI-R%20Technical%20Brochure.pdf) [<https://perma.cc/7746-SMAM>]. It is certainly in use in more states than any other risk assessment and may have more clinical research behind it. Yet, as reflected in the focus of this article and the relevant sources, LSI-R has received less media scrutiny and been less of a subject of public debate than COMPAS.

using data points that function like proxies for race.<sup>101</sup> Use of race-correlated factors, such as education level or employment status,<sup>102</sup> is not necessarily a problem. But continued use of race-correlated factors to predict recidivism can perpetuate the salience of these factors in predicting crime.<sup>103</sup> Additionally, many of these factors demonstrate predictive weight but are not related to the individual crime or the defendant's general criminal activity.

Critics may argue that it is not the role of sentencing risk assessment algorithms to ameliorate the upstream impacts of latent racial bias. But altering an algorithm is undoubtedly simpler and easier to implement than changing policing behaviors or generating new protocols for courts to engage with defendants. Certain remedies along this line of thinking have been proposed before.

#### E. Previously Suggested Solutions

As we have seen, there are a plethora of risk assessment algorithms used across the United States.<sup>104</sup> Yet, a robust body of scholarship proposes eliminating risk assessments from the sentencing process entirely. Professor Bernard Harcourt advances the argument that "risk" is simply a proxy for race.<sup>105</sup> Because bias is inherent in risk assessments, the only way to remove racial bias is to stop using risk assessments and instead employ alternative solutions.<sup>106</sup> This argument builds on Harcourt's previous work detailing the pitfalls of actuarial assessments at sentencing and has received other scholarly

---

101. See generally John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. CLINICAL PSYCHOL. 489, 499 (2016) (chronicling the growing interest in risk assessments and clarifying the roles risk assessments can play at sentencing). Monahan and Skeem are scrupulous in their use of the term "proxy." While they accept that criminal history is a proxy for risk, they refute the idea that criminal history is a proxy for race. Criminal history is more strongly correlated with recidivism than race is, leading criminal history to be the dominant factor. *Id.*

102. Michael Tonry, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 FED. SENT'G REP. 167, 167 (2014).

103. See Starr, *supra* note 34, at 808.

104. EPIC Algorithm List, *supra* note 14. COMPAS is used in five states and the LSI-R is used in seven states, while California uses an adapted version of Washington's LSI-R and many other states have adopted their own state-specific tools.

105. Harcourt, *supra* note 98, at 238.

106. *Id.* at 241. Harcourt proposes that in place of risk assessments we reduce sentence lengths, eliminate mandatory minimums, reduce sentences for drug-related crimes, and expand the use of alternative supervision programs.

support.<sup>107</sup> Yet, over the last decade we have actually seen a trend towards the increased use of risk assessment tools.<sup>108</sup> Why might this be?

Truthfully, the alternative solutions Harcourt suggests are difficult to implement.<sup>109</sup> Judges, legislators, and even private companies themselves can deflect blame onto an algorithm for any mistake, rather than justify their own decisions. As such, it is highly unlikely that we will see the elimination of risk assessments at the state or federal level anytime soon.

An equally vigorous subset of scholarship examines the viability of reducing the number of factors that risk assessment algorithms incorporate. By and large, the algorithms in question are protected under the law of trade secrets. While we can analyze the publicly available questionnaires and reverse-engineer crude representations of these algorithms through the publicly available data (like that of COMPAS in Broward County) there may be unknown quantities incorporated in these algorithms.<sup>110</sup> Additionally, we cannot know how algorithms weigh factors relative to each other. That said, we can still examine which factors are generally predictive of recidivism and winnow those down to the ones that are necessary.

Proponents of reducing the number of factors in risk assessment tools derive support, in part, from studies that exclude demographic and socioeconomic factors “without losing any significant predictive value.”<sup>111</sup> This phenomenon has borne out in reality, as certain risk assessment tools use significantly fewer variables than COMPAS. For example, the Virginia Criminal Sentencing Commission’s nonviolent risk assessment tool for larceny asks only five

---

107. HARCOURT, *supra* note 23, at 2; *see also* Starr, *supra* note 34, at 817 (alluding to Harcourt’s “thorough critique of the use of risk prediction in criminal justice” in *Against Prediction*); *Against Prediction - Review Quotes*, UNIV. CHI. PRESS (2007), <https://www.press.uchicago.edu/ucp/books/book/chicago/A/bo4101022.html> [<https://perma.cc/ZV8M-SR42>] (sharing the reviews of Malcolm Gladwell, David Mann and Peter Moskos, among other scholars).

108. HARCOURT, *supra* note 23, at 2 (“Most scholars, criminal justice practitioners, and public citizens embrace the turn to actuarial methods as a more efficient, rational, and wealth-maximizing tool to allocate scarce law enforcement resources.”).

109. Harcourt, *supra* note 98, at 241.

110. Larson et al., *supra* note 9.

111. Starr, *supra* note 34, at 851; *see* Joan Petersilia & Susan Turner, *Guideline-based Justice: Prediction and Racial Minorities*, in PREDICTION & CLASSIFICATION: CRIMINAL JUSTICE DECISION MAKING 151, 153–54, 160 (Don M. Gottfredson & Michael Tonry eds., 1987).

questions, while the number of variables increases for more serious crimes.<sup>112</sup> Recent studies also show that simpler tools can be just as predictive. Dressel and Farid, as one example, found that a classifier “based on only two features—age and total number of previous convictions—performs as well as COMPAS” in terms of predictive accuracy.<sup>113</sup>

But in all of these models, one factor remains constant: previous convictions. As the most predictive of recidivism relative to all other factors, this should come as no surprise.<sup>114</sup> However, number of previous convictions is also strongly correlated with race.<sup>115</sup> While eliminating other superfluous factors might achieve goals such as reducing resources or making risk assessments more accessible, the necessity of accounting for “previous convictions” jettisons the utility of this approach in solving the problem of racial bias hiding in our algorithms. Perhaps to dispel racially disparate impacts, we must directly grapple with race.

### III. ACCOUNTING FOR RACE: EQUALIZED ODDS-BASED RISK ASSESSMENTS

Section III.A of this Part will examine how race might be explicitly incorporated into risk assessments. Sections III.B and III.C will then propose basic solutions to the problem of racially disparate impacts in criminal sentencing that achieve the goals of affirmative action, even if not necessarily viewed as formal affirmative action programs. To conclude, Section III.D will offer a new proposal: a race-conscious, equalized odds-based risk assessment that could reduce the

---

112. VA. CRIM. SENT’G COMM’N, LARCENY WORKSHEET 6 (2018), [http://www.vcsc.virginia.gov/worksheets\\_2019/Larceny.pdf](http://www.vcsc.virginia.gov/worksheets_2019/Larceny.pdf) [<https://perma.cc/F35W-TU82>]. The factors accounted for in this worksheet are age, gender, prior convictions, prior incarcerations, and whether the individual was legally restrained at time of offense. Granted, this only applies to non-violent offenders, and it only determines whether an offender should receive alternative punishment (for low risk) or whether they should be incarcerated. For Sections A–C on this worksheet, the number of variables used is either 19 or 20.

113. Dressel & Farid, *supra* note 39, at 3; see also Jongbin Jung et al., *Simple Rules for Complex Decisions* 1 (Stan. Univ., Working Paper No. 1702.04690, 2017), <https://arxiv.org/pdf/1702.04690.pdf> [<https://perma.cc/QB9M-KY6C>] (presenting a method of statistics-based decision-making with rules that “take the form of a weighted checklist, can be applied mentally, and nonetheless rival the performance of modern machine learning algorithms”).

114. Starr, *supra* note 34, at 851 (citing Petersilia & Turner, *supra* note 111, at 171 fig.1).

115. Harcourt, *supra* note 98, at 239.

racial inequities currently created by predictive parity-based algorithms.

### A. Adding Race

Some scholars, to a less thorough degree, have advanced the idea of directly including race as a variable in risk assessment algorithms. J.C. Oleson, for instance, has proposed using race directly, though simply as another factor to increase a risk assessment's predictive ability.<sup>116</sup> But if equalized odds, rather than accuracy, is the main objective, we should consider directly accounting for race in a more serious capacity, such as adjusting the weight of race-correlated factors. Aziz Huq floats this idea but notes that it would face legal obstacles akin to affirmative action.<sup>117</sup> In the current political and legal climate, adding affirmative action policies to a new field is clearly not practicable.<sup>118</sup> But as recently as 2013, the Supreme Court upheld affirmative action policies,<sup>119</sup> and in the near future, pursuing affirmative action-like policies may once again appear viable. If and when that occurs, the inevitable legal obstacles may not seem so difficult to surmount. And in a strange way, the legal battles over affirmative action in higher education provide a neat roadmap for applying affirmative action principles to risk assessment algorithms used at sentencing.

Affirmative action is, undoubtedly, a loaded phrase.<sup>120</sup> Depending on one's political and social views it may even sound radical. At the very least, it has been a subject of fierce debate in this country for decades. But at its core, the idea of affirmative action is to offset historical disadvantages of a minority group by altering a decision-making process to be more favorable towards that group. Generally, this is done by incorporating that group identity as a positive factor in the decision-making process and rebalancing the

---

116. J.C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329, 1399–402 (2011).

117. Huq, *supra* note 50, at 6.

118. Barocas & Selbst, *supra* note 92, at 715.

119. See, e.g., Fisher v. Univ. of Texas at Austin, 570 U.S. 297, 310–12 (2013) (holding that the Fifth Circuit had not correctly applied strict scrutiny review to the affirmative action program in use at the University of Texas).

120. See JOHN DAVID SKRENTNY, THE IRONIES OF AFFIRMATIVE ACTION: POLITICS, CULTURE AND JUSTICE IN AMERICA 6 (1996); see also Martha S. West, *The Historical Roots of Affirmative Action*, 10 LA RAZA L.J. 607, 607 (1998) (describing affirmative action as "a politically-loaded word").

weight of the other factors that are typically considered.<sup>121</sup> Though affirmative action is often conceived of as a formal program in higher education or the workplace, the core tenet of “leveling the playing field” can manifest in a variety of ways. Though the first two proposals might not read like affirmative action programs at first blush, all of the accompanying solutions are likely to have an effect on sentencing similar to the effect of affirmative action in higher education.

### B. Weighting Low Risk Scores Greater than High Risk Scores

As applied to risk assessments at sentencing, the simplest form of “affirmative action” has already been proposed. The Council to the Members of the American Law Institute has argued that low risk scores should be accorded more weight at sentencing than high risk scores, as low risk scores are more often accurate than high risk scores.<sup>122</sup> Granted, this suggestion seems to have been motivated by concerns about accuracy and resource scarcity, rather than concerns about racial inequality.<sup>123</sup> Nonetheless, in a manner opposite that of COMPAS, prioritizing the accuracy of low risk scores would likely have positive externalities along racial lines. Such an approach would disproportionately benefit black defendants with high risk scores (as we have seen, this group has a higher percentage chance of being misclassified). Discounting high risk scores would in turn create a downstream effect, leading judges to assign less lengthy sentences based on high risk scores, thereby ameliorating racial disparities at sentencing.

---

121. Accounting for race in a constitutionally acceptable way has proven difficult for many institutions. One example of how it is done in practice is by using race as a “plus factor.” But “plus factor” can be a misleading term, in that affirmative action systems do not distill race into a singular factor with a set weight relative to other factors. Rather, affirmative action admissions systems, like Harvard’s, consider race in the context of “a candidate’s life experience and background in addition to grades and test scores.” Courtney Rozen, *How Americans Feel About Affirmative Action in Higher Education*, NPR (Nov. 1, 2018), <https://www.npr.org/2018/11/01/658960740/how-americans-feel-about-affirmative-action-in-higher-education> [https://perma.cc/66XG-C7NL].

122. MPC SENTENCING DRAFT 2 (2011), *supra* note 95, at 56 (noting that “[f]rom an actuarial perspective, attempts to identify persons of low recidivism risk are more often successful than attempts to identify persons who are unusually dangerous”).

123. *See id.* The report goes on to note that moving low risk offenders out of prisons and into community service programs will conserve prison resources and reduce overall costs. *Id.* § 6B.09(d).

### C. Ensuring Validation of Algorithms for Race

Another relatively simple method of affirmative action can be applied in the context of validating risk assessment algorithms for race. As previously mentioned, a disturbingly low number of states have made public whether (and if so, how) their risk assessment algorithms have been validated.<sup>124</sup> The widely used LSI-R, for example, was developed and initially validated using only statistics from “Canadian offenders of predominantly Caucasian ethnic heritage.”<sup>125</sup> Similarly, later validation on samples of U.S. offenders was also conducted using data mostly from Caucasian offenders.<sup>126</sup> The repercussions of this, while not fully understood, showed the LSI-R had lower validity on black and Latino subjects than on Caucasian subjects.<sup>127</sup> Requiring risk assessment algorithms to be developed and tested using data from all ethnic and racial backgrounds would expose the blind spots of these algorithms, and either force their creators to adjust accordingly or provide courts with a legitimate reason to discount that risk assessment’s validity.

### D. A More Radical Solution: Affirmative Action for Sentencing Algorithms

Finally, this Note proposes a new solution: pursuing equalized odds in risk assessment algorithms, following the traditional mold of affirmative action. The college admissions and employment-seeking processes may seem a world away from criminal sentencing, but when viewed as critical junctures in the trajectory of a person’s life, they appear more reasonable bedfellows. Data points on a college application such as grades, test scores, and work experience provide a great deal of information to an admissions officer about an applicant’s likelihood of future success in higher education. But if any of these data points appear undesirable (e.g. academic or legal issues), this

---

124. *EPIC Algorithm List*, *supra* note 14.

125. Melinda D. Schlager & David J. Simourd, *Validity of the Level of Service Inventory-Revised (LSI-R) Among African American and Hispanic Male Offenders*, 34 CRIM. JUST. & BEHAV. 545, 546 (2007).

126. *Id.* Though Schlager and Simourd reported on studies done as of 2007, their findings appear to have held true at least through 2013, at the publication time of a student note citing this data. See McGarraugh, *supra* note 14, at 1097. This author has found no recent evidence to the contrary.

127. Compared with a study done by Andrews and Bonta in 1995, Schlager and Simourd’s study focusing on black and Latino male offenders showed lower internal consistency rates. Schlager & Simourd, *supra* note 125, at 553 (citing D.A. ANDREWS & J. BONTA, THE LEVEL OF SERVICE INVENTORY—REVISED (1995)).

information is often insufficient to explain the reasons for an applicant's struggles. For black applicants specifically, the admissions process fails to adequately capture an individual's experience with institutional racism. To fill this gap, universities across the country incorporate race as a "plus factor" in the admissions process.<sup>128</sup>

At present, the same approach is not taken in criminal sentencing. Yet, sentencing is the last stage in a criminal justice process that can be riddled with historic and systemic inequality. The causes of potential injustices, woven through arrest, bail, plea bargaining and trial, are often difficult to pinpoint.<sup>129</sup> Furthermore, certain factors may begin influencing defendants well before they ever personally interact with the court system. A risk assessment cannot account for all of these factors or their interplay, and the scope of institutional racism is often too difficult for a judge to grasp, weigh, and act upon. In this way, sentencing procedure fails black defendants in the same way that admissions processes prior to affirmative action failed black applicants.

With a risk assessment like COMPAS, race could easily be considered as a plus-factor for a lower risk score in the algorithm. By explicitly acknowledging a defendant's race, the algorithm could then identify a number of static factors highly correlated with race but unrelated to a defendant's criminal activity and discount the weight of these factors in the overall risk score.<sup>130</sup> More specifically, it could

---

128. "Plus factor" can be a misleading term, in that affirmative action systems do not distill race into a singular factor with a set weight relative to other factors. Rather, affirmative action admissions systems, like Harvard's, consider race in the context of "a candidate's life experience and background in addition to grades and test scores." Rozen, *supra* note 121.

129. There are certainly injustices that are possible to pinpoint and quantify. But some important questions have yet to be answered. For example, is there a greater police presence in a predominantly black community because it has a high crime rate, or vice-versa? Did the defendant plead guilty because he was guilty or because he could only afford an inexperienced lawyer?

130. One static factor might be the level of "family disruption" as indicated by the answers to Questions 55 and 56 of the Sample COMPAS Risk Assessment Questionnaire: "How often have you moved in the last twelve months?" and "Do you have a regular living situation?" A defendant without a stable community or support system may be statistically more likely to recidivate. However, we also know that black families are more likely to experience "family disruption." Sampson & Lauritsen, *supra* note 85, at 335; HAWKINS ET AL., *supra* note 86, at 2. Logic suggests that a defendant dealing with "family disruption" would more likely benefit from alternative measures that focus on community-building and rehabilitation than from incarceration, which only serves to exacerbate family

identify factors that indicate a higher likelihood of rehabilitation, as has recently been proposed in a report from the Congressional Research Service.<sup>131</sup> Finally, creators of a race-conscious risk assessment could adjust the weight of race as a “plus factor” as time passes if the data shows that racial biases are being filtered out of the criminal justice system.

## CONCLUSION

Algorithmic risk assessments currently used in criminal sentencing, such as COMPAS, may be fair in one sense. But as we have seen, even a reasonable, widely accepted definition of fairness can lead to patently unfair disparate impacts falling along racial lines. Moreover, an algorithm built by humans, using past data indirectly infused with decades of racial bias from American society, can reflect and reinforce our society’s prejudices. Predictive accuracy is certainly an important value for both legislatures and courts. Employing algorithmic risk assessments that satisfy predictive parity can adequately fulfill our commitment to that value. And yet, the founding ideals of this nation remind us of the moral imperative to equalize the odds in our criminal justice system.<sup>132</sup> We should not dismiss this definition of fairness because it is inconvenient or difficult to implement.

The best course of action, then, is to develop race-conscious risk assessment algorithms for criminal sentencing. Though the general framework of affirmative action has not yet been applied to sentencing and algorithmic fairness, the work of scholars in this field supports the underlying logic of that framework and can be seen as building to this

---

disruption. The algorithm could thus flag this factor as an indicator for alternative rehabilitative measures and reduce the coefficient (weight) of this data point in assessing risk. Another factor applicable here might be the answer to Question 67: “In your neighborhood, have some of your friends or family been crime victims?” As personal experiences with murder and robbery are also correlated with “family disruption,” if a defendant answers “Yes” to this question, the algorithm could reduce the coefficient of this data point in assessing risk. NORTHPOINTE INC., SAMPLE COMPAS RISK ASSESSMENT 5 (2011), <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html> [https://perma.cc/66FJ-5HF9].

131. See NATHAN JAMES, CONG. RESEARCH SERV., R44087, RISK AND NEEDS ASSESSMENT SYSTEM IN THE FEDERAL PRISON SYSTEM 2 (July 2018) (“‘Criminogenic needs,’ are factors that contribute to criminal behavior that can be changed and/or addressed through interventions.”).

132. See *supra* note 96 and accompanying text.

kind of solution. Moreover, there is space in the law for such an approach to work. While the current political environment may be inhospitable to any type of affirmative action proposal, adopting a more equitable definition of fairness for criminal sentencing could be viable someday soon—perhaps when we as a society more fully grasp the implications of algorithms helping dispense justice.