Compounding Injustice: The Cascading Effect of Algorithmic Bias in Risk Assessments

TIM O'BRIEN*

ABSTRACT

The increasing pervasiveness of algorithmic tools in criminal justice has led to an increase in research, legal scholarship, and escalating scrutiny of automated approaches to consequential decisionmaking. A key element of examination in literature focuses on racial bias in algorithmic risk assessment tools and the correlation to higher likelihoods of high bail amounts and/or pretrial detention. These two phenomena combine to initiate a cascading effect of increased likelihoods for conviction, incarceration, harsher sentencing, higher custody levels, and barriers to parole, leading to negative impacts on other factors unrelated to criminal history, all of which feed into subsequent assessment instruments for defendants who are re-arrested. This escalating cascade of algorithmic bias errors has particularly dire consequences for Black defendants, who are statistically more likely to receive higher failure-to-appear (FTA) and recidivism risk scores than white defendants, and who are thus more likely to be negatively impacted by subsequent decisions, both human and computer-aided, throughout the criminal justice process. This is periodically referred to in literature as 'disparate impact' but lacks a deeper examination of broad-based effects. This Article endeavors to advance that examination by looking across multiple elements of criminal procedure and beyond to aid in understanding cascading effects and consequent injustices suffered by Black defendants due to the continued automation and encoding of societal biases into the criminal justice process.

TABLE OF CONTENTS

INTRODUCTION 41

I. CASCADING EFFECTS 44

A. Proxy Domains 45

* Master of Jurisprudence Candidate at the University of Washington School of Law, studying cyberlaw and public policy; MBA, Northwestern University Kellogg School of Management; BS Engineering, Purdue University. Tim leads Ethical AI Advocacy at Microsoft and is responsible for programs to drive and promote responsible development and use of technology, inclusive of public policy and the ethics of artificial intelligence. Tim is also a Guest Lecturer at INSEAD, one of the world's leading business schools, where he speaks on cross-cultural leadership and ethics. He can be contacted at tobrien@microsoft.com or at timo66@uw.edu. The author wishes to thank Ryan Calo for his supervision of this project and Vincent Southerland for his valuable insights on cascading effects. The author is also grateful to Helen Anderson, Elizabeth Bender, Alexandra Chouldechova, Cynthia Conti-Cook, Ece Kamar, Kristian Lum, and Hanna Wallach for their knowledge and guidance throughout the process of researching and writing this Article.

B. Chaos Theory and the Law 48

II. ALGORITHMIC ASSESSMENT 49

A. Four Generations of Risk Assessment 50

B. Applications 52

1. Pretrial 52

2. Sentencing 53

3. Classification 54

4. Reentry and Parole 55

III. THE BIAS PROBLEM 56

A. ProPublica and COMPAS 58

B. Beyond ProPublica 60

IV. IMPACT ON CRIMINAL JUSTICE 62

A. Assessment, Detention, and Conviction 62

B. Sentencing and Classification 65

C. Incarceration and Parole 66

D. Reentry 69

E. Re-Arrested, Re-Assessed 72

V. FUTURE DIRECTIONS: RESEARCH AND POLICY 74

A. Towards a Research Agenda 74

1. Causal Linkage 75

2. Amplification and Compounding 76

3. Judicial Reliance 77

B. Policy Interventions 78

1. Opening the Black Box 78

2. Revising Rule Exceptions 79

3. Instituting Checks 81

CONCLUSION 81

INTRODUCTION

When 19-year-old Terrence Wilkerson was arrested in 2000 for a robbery he did not commit, he was judged a high risk for failure-to-appear (FTA), resulting in a large bail amount.[1] In describing his arraignment, Wilkerson notes, "I'm African American, I have braids, I look a certain way."[2] Unable to post bail, he was jailed before accepting a plea agreement that sent him to state prison for two years. He explains his decision, saying, "I took a plea for something I didn't do. I'd already spent 10 months on Rikers Island . . . I knew I had to take that plea. A few more years upstate? At least I'll get away from Rikers Island."[3]

When Wilkerson was arrested 17 years later, again for a robbery he did not commit, he was subjected to an algorithmic risk assessment[4] that considered his prior "conviction," and his bail was set at $25,000. Fortunately, the judge considered the lack of evidence and reduced bail to $2,500, after which Wilkerson was released and able to work with his lawyers, resulting in an acquittal. Terrence Wilkerson's experience raises questions about complex interactions between the algorithmic prediction of FTA (and recidivism) risk and bail amounts, the presence of racial bias in these predictions, and the downstream impact of bias. More broadly, these dynamics pose additional questions about greater likelihoods for pretrial detention, the impact of pretrial detention on case disposition, the effects of severed ties to family, job, education, and social stability, and the broader role of algorithms, formulas, and mathematics in shaping the criminal justice process.

The use of mathematics in administering justice gained attention in 1971, when Laurence Tribe published a seminal article in the *Harvard Law Review*, *Trial by Mathematics*,[5] that considered the accuracy and appropriateness of mathematical methods and statistical probabilities in civil and criminal trials. The article was inspired in part by *People v. Collins*,[6] a 1968 robbery case in which the prosecutor's inability to positively identify the interracial couple on trial led him to summon a math professor from a local college to testify to the low probability that they *did not* commit the crime. The jury returned a guilty verdict, only to be overturned on appeal by the California Supreme Court, which criticized the use of statistical reasoning, asserting that "[m]athematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not cast a spell over him."[7] While the context of the court's holding pertains to the trial itself, the

1. Elizabeth Bender, Kristian Lum & Terrence Wilkerson, *FAT* 2018 Translation Tutorial: Understanding the Context and Consequences of Pre-trial Detention*, YOUTUBE (Apr. 18, 2018), https://www.youtube.com/watch?v=hEThGT-_5ho (recording of panel at FAT* 2018 Conference).

2. *Id.*

3. *Id.*

4. In New York City, pretrial release assessment is administered by the New York Criminal Justice Agency (CJA) using an algorithmic assessment instrument. *See* N.Y. CRIM. JUST. AGENCY, *NYCJA Release Assessment*, https://www.nycja.org/release-assessment (last visited Jan. 21, 2021) [https://perma.cc/2NMK-GL9D].

5. Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

6. People v. Collins, 438 P.2d 33 (Cal. 1968).

7. *Id.* at 33.

increasing pervasiveness of algorithmic decision-support tools in the criminal justice process challenges this sentiment.

There has been a great deal of research in recent years on the "spell" of mathematical formulas and the presence of racial bias in algorithmic risk assessment instruments, primarily in the context of pretrial decisions concerning FTA and recidivism, and secondarily in the context of sentencing. But the problem is much worse in aggregate. As this Article will show, if a person is arrested, convicted, classified, incarcerated, and paroled, then by the time they re-enter society, they have been subjected to a minimum of a half-dozen consequential decisions that were either made or influenced by algorithms. If the defendant is Black, they have been consequently exposed to the cascading effect of algorithmic bias errors, stacked on top of one another and magnified, during their journey through the criminal justice process.[8] Furthermore, these cascading effects continue into societal reentry, with negative impacts on employment, housing stability, family life, physical and mental health, and likelihood of getting re-arrested. If a Black person is re-arrested, the cascading effects pick up where they left off, with a biased pretrial assessment followed by an escalating statistical propagation pattern that is, for all practical purposes, irreversible.

This decisionmaking pattern is distinguished by the subsequent impact of what are seemingly small inputs in the early stages of the process, such as answers to assessment survey questions immediately following arrest. In 1972, Edward Norton Lorenz presented a paper to the Association for the Advancement of Science entitled, "*Predictability: Does the Flap of a Butterfly's Wings in Brazil set off a Tornado in Texas?*"[9] In what would become the branch of mathematics known as "chaos theory," Lorenz described complex, dynamic systems in which "tiny changes to the initial conditions can lead to widely diverging outcomes."[10] The so-called "butterfly effect" is generally accompanied by recursion: taking the output of one prediction in a system and applying it as an input to the next. The cascading effects of algorithmic bias in assessment instruments are emblematic of the butterfly effect, in which "initial conditions" in the form of answers to instrument survey questions that seem small in isolation invariably lead to significant compounding errors in the sequence of decisions that progress over the course of the criminal justice process. As criminal justice increasingly relies on algorithmic tools for decision support, it creates a butterfly

8. This binary view of racial stratification in the criminal justice process warrants a note on terminology. In a caveat to her article on bias in risk assessment, Mayson (2019) laments references to race "in the crass terminology of 'black' and 'white.' This language reduces a deeply fraught and complex social phenomenon to an artificial binary." Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2226 (2019). Huq (2019) shares a similar caveat: "[a] focus on a black-white binary is warranted here as a way of clarifying the fundamental conceptual stakes. It is obviously inadequate as a general account of racial equity in policing, and I do not intend it as such." Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1103 (2019). I share this sentiment, only adopting this coarse-grained partitioning of groups of people for purposes of simplicity and consistency with existing literature. Ideally, we will move beyond this taxonomy and into one that better reflects the complexity of racial makeup in America.

9. Edward Lorenz, *Predictability: Does the Flap of a Butterfly's Wings in Brazil set off a Tornado in Texas?*, Association for the Advancement of Science 139th Meeting (1972).

10. MATH VAULT, *The Definitive Glossary of Higher Mathematical Jargon, Chaos*, https://mathvault.ca/math-glossary/#chaos (last visited Jan. 21, 2021) [https://perma.cc/HS97-WK5C].

effect of racial bias, with statistical errors in one domain (e.g., pretrial detention) magnified and ingested by the next domain (e.g., sentencing), followed by further compounding (e.g., classification), followed by even more algorithmic assessments for everything from prison misconduct predictions to parole decisions.

This Article will provide a better understanding of cascading effects by connecting the demonstrable existence of racial bias in algorithmic tools with existing literature on downstream impact, thus placing both of these observed phenomena into a broader context. Downstream impact is often referred to in literature as "disparate impact" and is used in different contexts with different meanings. Chouldechova (2017)[11] and Mayson (2019)[12] both reference disparate impact to place a focus on the fairness of decisionmaking outputs from a given algorithm, thus narrowly focused on a point in time defined by a procedural decision. Chouldechova (2017) goes on to assert disparate impact is a "social and ethical concept, not a statistical concept."[13] Huq (2019) makes reference to "spillover effects," writing that these effects "for Black families and communities appear to be larger in magnitude than the spillover effects in white communities, even controlling for the extent of coercion."[14] There is a large body of research that isolates specific elements of ensuing impact and provides deeper understanding of each element, inclusive of statistical correlations to earlier procedural decisions, many of which have been shown to include a racial bias. This Article will provide a synthesis of this body of work to frame the broader context of cascading effects on Black defendants.

For purposes of this Article, we will restrict the discussion to exclude predictive policing and begin with risk assessment algorithms at the point of arrest through incarceration and reentry. We will consider the cascading effects of racial bias in the initial FTA and recidivism assessments on both subsequent assessments in the criminal justice process and following reentry into society and re-integration into family, social networks, and the work environment.

Section I will focus on error propagation and cascading effects by considering proxy domains—that is, fields other than criminal justice and sentencing—in which error compounding, domino effects, and propagation patterns have been observed and studied, as well as chaos theory as a proxy to understand these complex interactions. Section II will focus on the assessment instruments themselves and the history behind their pervasive presence, including the main elements of the criminal justice process in which algorithms play a role and the progression of technology advances that motivates their increasing adoption. Section III will discuss the well-researched problem of racial bias in algorithmic assessment

11. Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 BIG DATA 153, 153-54 (2017).

12. Mayson, *supra* note 8, at 2218.

13. Chouldechova, *supra* note 11, at 154.

14. Huq, *supra* note 8, at 1105.

instruments, with particular attention to ProPublica's 2016 research[15] and the ongoing debate centered on tension between definitions of "fairness" and "accuracy." Section IV will expand the investigation of cascading effects by connecting research on racial bias in algorithmic instruments to statistical correlations between decisions in criminal procedure and their ensuing impact, including post-release effects. Section V will propose an agenda for follow-on research and describe concrete steps available to policymakers and legislators to provide interventions to mitigate the cascading effects that play out every day in the U.S. criminal justice system.

I. CASCADING EFFECTS

The cascading effects of bias that compound and propagate through a system or process begin with the accuracy of a given algorithmic prediction. For algorithms that predict the likelihood that individual data points will fall into predetermined categories, accuracy is calculated as the sum of correct predictions.[16]

But these systems rarely exist in isolation—they are often part of a process in which the outputs of one algorithm either indirectly influence or provide the inputs to the next algorithm. If any algorithm in a sequence has a systematic error caused by biases in data, then the net result is obviously biased. But if more than one algorithm in the process is producing systematic errors, then those errors are compounded as the sequence unfolds in a linear fashion, often with no feedback loop for correction of errors. Empirical research into cascading effects in the criminal justice domain is lacking. Hellman (2017) describes "compounding injustice" as an effect of "indirect discrimination," i.e., the disparate impact of "an action that exacerbates the harm caused by the prior injustice because it entrenches the harm or carries it into another domain."[17] This characterization is principally correct but falls short of addressing the compounding of systematic errors in the statistical or algorithmic sense.

Because, to a growing extent, the conveyance of harm to multiple domains within the criminal justice process occurs algorithmically, this Article takes a generalized view of cascading effects by looking to other domains in which the phenomenon of compounding errors and cascading effects have been researched and studied. We can utilize these domains as proxies to draw inferences about how systematic errors in assessment algorithms compound as they make their way through the criminal justice process.

---

15. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/DR3W-RTKX].

16. "Accuracy" in this context refers to the closeness of a measured value to a known value. In binary classification, a confusion matrix (or error matrix) is a two-by-two table that describes the rate of true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP). Accuracy = (TP+TN)/(TP+FP+TN+FN). *See generally* Kurtis Pykes, *Confusion Matrix "Un-confused": Breaking down the confusion matrix*, TOWARDS DATA SCI. (Feb. 16, 2020), https://towardsdatascience.com/confusion-matrix-un-confused-1ba98dee0d7f [https://perma.cc/4998-7V5Z].

17. Deborah Hellman, *Indirect Discrimination and the Duty to Avoid Compounding Injustice* (Va. Pub. L. & Legal Theory, Research Paper No. 2017-53, 2017).

A. Proxy Domains

In the domain of employment law, Kessler (2017) looks at categories of employment discrimination such as disparate treatment, disparate impact, and sexual harassment. She describes these phenomena as "compartmentalized," and thus not reflective of the reality that they are often experienced concurrently, arguing that "various types of exclusion often add up to significant inequalities, even though seemingly insignificant when considered in isolation."[18] The resultant cascading effect stems from the interactions among variables with complex cross-correlations, in which "worker inequality often results from a series of discriminatory conditions or triggers that combine and interact in ways that, over time, may lead to large differences in employee status and pay due to their cumulative and mutually reinforcing nature."[19] This discriminatory cascade can begin with something as innocuous as a name, as O'Neil (2017) describes, recounting a research study in which fake résumés were submitted for job openings, with each résumé having comparable qualifications, but modeled for race: "[h]alf featured typically white names like Emily Walsh and Brendan Baker, while the others with similar qualifications carried names like Lakisha Washington and Jamal Jones, which would sound African American. The researchers found that white names got fifty percent more callbacks than the Black ones."[20] These human prejudices appear as patterns in data, which are then defined mathematically, expressed as rules in the form of algorithms, and thus embedded into algorithmic decision systems. A biased outcome will result, triggering subsequent discriminatory outcomes that are often non-obvious in isolation but accumulate over time. As Kessler (2017) writes, "by failing to recognize the dynamic, interactive processes . . . legal and political discourses on discrimination mask the pervasive and powerful role of institutions in creating inequality."[21] The institution that is the criminal justice system has a number of parallels to workplace discrimination, but none more insidious than the increasingly pervasive role of algorithms and their dynamic dependence on each other.

In the world of global finance, statistical methods are employed to study market fluctuations and cross-correlations among variables. Unlike the criminal justice system, financial markets are subject to assumed randomness, but nonetheless offer a partial understanding of the complex mathematical interactions between attributes and resultant cascading effects, often described in literature as "financial contagion."[22] Degryse and Nguyen (2004) describe how "[c]ontagion on interbank

18. Laura T. Kessler, *Employment Discrimination and the Domino Effect*, 44 FLA. ST. U. L. REV. 1041, 1041 (2017).

19. *Id.*

20. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 113 (2016).

21. Kessler, *supra* note 18.

22. Rudiger Dornbusch, Yung Chul Park & Stijn Claessens, *Contagion: Understanding How It Spreads*, 15 WORLD BANK RSCH. OBSERVER 177 (2000).

markets can occur . . . when the collapse of a bank induces a domino effect."[23] The article goes on to explain the process through which:

> The failure of one individual bank may initiate a domino effect if the non-repayment of interbank obligations by the failing bank jeopardizes the ability of its creditor banks to meet their obligations to their (interbank) creditors. Contagion occurs then 'mechanically' through the direct interlinkages between banks. Domino effects may arise across regions or bank types.[24]

In this context, "mechanically" is a partial reference to the computerized, algorithmic systems characteristic of modern banking and the digital connections required to process transactions at scale. In the midst of the global financial crisis,[25] Degryse, Elahi, and Penas (2010) used aggregate cross-border liability risk exposures to further describe how non-repayment of foreign debts "starts a domino effect that impacts other banking systems worldwide."[26] But the impact of the financial crisis was felt far beyond the banking system, with domino effects propagating through society and the lives of individuals. The impact on American cities is described by Dickerson (2009) as "a devastating effect . . . because of the lower [tax] revenues . . . due to properties which are now less valuable."[27] The domino effect is directly traceable to reductions in municipal services, urban blight due to home vacancies, increases in violent crime, arsons committed by homeowners in financial distress, and homelessness, all of which place a further strain on police departments and social services.[28]

In healthcare, cascading effects have been widely studied in regard to specific diagnoses and courses of treatment involving narrowly defined diseases and medical conditions, all involving complex sets of interdependent variables and patient attributes.

For purposes of this Article, however, the most relevant analog is the occurrence of medical error,[29] which (excluding COVID-19) is the third most common cause of death in the United States.[30] Coughlan, Powell, and Higgins (2017) describe adverse outcomes in pregnancy, labor, and delivery as having "a domino effect with three groups being involved – the patient (first victim), the staff (second victims) and the

23. Hans Degryse & Giang P. Nguyen, *Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System* (Nat'l Bank of Belgium, Working Paper No. 43, 2004).

24. *Id.*

25. Also referred to as the "Financial crisis of 2007-08" and the "Subprime Mortgage Crisis." *See* Brian Duignan, *Financial crisis of 2007–08*, ENCYCLOPEDIA BRITANNICA (2019), https://www.britannica.com/event/financial-crisis-of-2007-2008 [https://perma.cc/NKW4-D4QC].

26. Hans Degryse, Muhammad Ather Elahi & Maria Fabiana Penas, *Cross-Border Exposures and Financial Contagion*, 10 INT'L REV. FIN. 209, 210 (2010).

27. A. Mechele Dickerson, *Over-Indebtedness, the Subprime Mortgage Crisis, and the Effect on U.S. Cities*, 36 FORDHAM URB. L.J. 395, 417 (2009).

28. *Id.* at 418.

29. Maité Garrouste-Orgeas et al., *Overview of Medical Errors and Adverse Events*, 2 ANNALS INTENSIVE CARE 2, 2 (2012).

30. JOHNS HOPKINS MED., *Study Suggests Medical Errors Now Third Leading Cause of Death in the U.S.* (May 3, 2016), https://www.hopkinsmedicine.org/news/media/releases/study_suggests_medical_errors_now_third_leading_cause_of_death_in_the_us [https://perma.cc/3ZN4-4WSE].

organization (third victims)."[31] Ellahham (2018) takes an even broader view with regard to medical error, writing:

> The entire health care system works as a single entity; hence, if one component falls, all the others fall, creating a domino effect. Identifying the vulnerable component of the health care framework and providing adequate support to it can possibly break this cycle and protect other members from the possible collapse.[32]

The study of these effects entails an understanding of the social science and human psychology behind the impact a medical error has on all actors in the healthcare system: the patient and his/her family, the healthcare professional who committed the error, the hospital support and administrative staff, the healthcare organization and its corresponding reputation, and the healthcare community generally, whose very existence relies on the public trust.[33] This broader view is instructional, as the net effect of bias in criminal justice has a similar impact radius on friends, family, employers, and socioeconomic status, all of which are subjects of pretrial risk assessment survey questions in the event an offender is re-arrested.

Domino effects are also studied widely in safety engineering, with particular attention given to risk modeling, frequency estimation, and propagation patterns of accident chains. Like financial markets, the triggering catalyst in industrial accidents, the "primary event," is often what Taleb (2007) calls a "Black Swan," a statistical outlier that is rare, improbable, and not reasonably foreseen.[34] Regardless, the propagation pattern of the resulting domino effect can be modeled using what Khakzad, Khan, Amyotte, and Cozzani (2013) describe as "units" in process engineering, writing:

> [P]otential secondary units are those adjacent units that are more likely to contribute to the domino effect. The inclusion of secondary units in the domino effect not only intensifies the accident, causing more severe consequences, but also helps the domino effect escalate to the next level by impacting tertiary units. The escalation vectors originating from secondary events in turn trigger other accidents.[35]

With respect to bias in risk assessment instruments, a pretrial detention is analogous to a "primary event, and "units" are analogous to the decisions in criminal procedure that follow. "Escalation vectors" are simply the presence of bias in subsequent decisions that serve to extend the net effect far beyond parole and reentry into society.

---

31. Barbara Coughlan, Doreen Powell & Mary F. Higgins, *The Second Victim: A Review*, 213 EUR. J. OBSTETRICS & GYNECOLOGY & REPRO. BIOLOGY 11, 11 (2017).

32. Samer Ellahham, *The Domino Effect of Medical Errors*, 34 AM. J. MED. QUALITY 412, 412 (2019).

33. Coughlan et al., *supra* note 31.

34. *See* NASSIM NICHOLAS TALEB, THE BLACK SWAN: THE IMPACT OF THE HIGHLY IMPROBABLE (2007).

35. Nima Khakzad, Faisal Khan, Paul Amyotte & Valerio Cozzani, *Domino Effect Analysis Using Bayesian Networks*, 33 RISK ANALYSIS 292, 294 (2013).

B. Chaos Theory and the Law

Lorenz (1972) pioneered chaos theory in the course of studying mathematical modeling of weather patterns,[36] but the field of study quickly expanded beyond meteorology and into virtually every other branch of science. The systems studied in chaos theory are deterministic: given a known initial state, one can theoretically predict the future state. Risk assessment instruments are technically probabilistic models, given their focus on likelihoods, in that they classify offenders by matching them with known groups of offenders with similar attributes. That said, the evolving approach to assessing recidivism risk has over time placed more emphasis on a more precise measurement of "initial state," as evidenced by, for example, the 137-question COMPAS pretrial assessment survey.[37] One could conclude that, in practice, modern risk assessment instruments are deterministic: a fixed set of inputs will produce the same outputs with every iteration, but a small change to inputs can lead to a significantly changed output over time.

Application of chaos theory to the legal domain exists in literature, building on Tribe's (1989) advocacy for use of analogies from the physical sciences in understanding the law.[38] Scott (1993) invokes chaos theory as a proposal to resolve contradictions and disorder in the legal system, writing, "even when we understand interactions very well, and even when the applicable laws are quite accurate and clear, results in specific cases still can be impossible to predict—even though recurring patterns are discernable and remarkably durable."[39] The article goes on to assert, "[b]y explicitly applying [chaos theory] to law, it becomes clear that even slight differences in the facts of cases result in wildly disparate judicial outcomes."[40] Hayes (1992) provides a view on applying the theory in the legal domain, writing "chaos is essential to understand law, and jurisprudence ignores chaos at its risk."[41] Regarding risk assessment, if one considers a classification scale as a set of "smooth edges" that exist between high, medium, and low risk, "chaos theory shows that . . . just as fractal shapes have complex borders rather than smooth edges, legal rules and doctrines also require careful examination to determine on which side of the jagged line a specific case lies."[42] Possibly foreshadowing the law's increasing reliance on scientific method to predict outcomes, Reynolds (1991) suggests the legal profession can overreach, writing:

> [L]awyers generally demand more from our theories than do scientists nowadays; we try too hard to find theories that predict outcomes, and we despair unnecessarily

36. Lorenz, *supra* note 9.

37. NORTHPOINTE INC., *Sample Risk Assessment COMPAS* [hereinafter *COMPAS Survey*], https://assets.documentcloud.org/documents/2702103/Sample-Risk-Assessment-COMPAS-CORE.pdf (last visited Jan. 21, 2021) [https://perma.cc/22ZF-QBTL].

38. Laurence H. Tribe, *The Curvature of Constitutional Space: What Lawyers Can Learn from Modern Physics*, 103 HARV. L. REV. 1, 1 (1989).

39. Robert E. Scott, *Chaos Theory and the Justice Paradox*, 35 WM. & MARY L. REV. 329, 331 (1993).

40. *Id.* at 348.

41. Andrew W. Hayes, *An Introduction to Chaos and Law*, 60 UMKC L. REV. 751, 752 (1992).

42. *Id.* at 766.

> when such efforts fail. Worse yet, we do these things largely out of a misguided effort to be 'scientific,' when scientists themselves have managed to come to terms with uncertainty, and even to put it to work.[43]

Algorithmic risk assessments have two key attributes that ring familiar to students of chaos theory: extreme sensitivity to small changes to initial conditions (e.g., survey answers), and recursion (i.e., taking the result of one decision and applying it to the next). For Black defendants in the criminal justice system, these two attributes are constant companions, as assessment surveys are administered with bias encoded in the instrument, and the algorithmic predictions that result set in motion a cascading, irreversible sequence of events that serve to magnify the effects of bias.

II. ALGORITHMIC ASSESSMENT

Consequential decisions in criminal justice are increasingly subject to influence by algorithmic decisionmaking and decision-support technologies, inclusive of risk assessment tools. The advent of risk assessment tools in recent decades needs to be considered in the context of policy changes and public attitudes toward crime and punishment that have invariably led to mass incarceration in America.

The early 1970's marked the era in which the prison system in America moved from a rehabilitative system to a punitive one.[44] The catalyst was the "War on Drugs,"[45] as it was termed by President Richard Nixon in 1971, and his declaration that drug abuse was "public enemy number one."[46] Legislation such as the Comprehensive Drug Abuse Prevention and Control Act of 1970[47] sought to address (among other things) the local distribution and possession of controlled substances. By 1973, there were over 300,000 arrests reported by the FBI's Uniform Crime Reports (UCR) for drug law violations, out of a total 9 million arrests nationwide for all offenses.[48] Thus began an exponential growth of incarcerated Americans, inclusive of inmates in state prisons, local jails, and federal prisons, with incarceration rates highest in the South and lowest in the Northeast.[49] This context is important to consider in any discussion of history's evolving views and tactics employed to differentiate lower-risk offenders from higher-risk offenders and the role of immutable factors

43. Glenn Harlan Reynolds, *Chaos and the Court*, 91 COLUM. L. REV. 110, 111 (1991).

44. "Until the mid-1970s, rehabilitation was a key part of U.S. prison policy. Prisoners were encouraged to develop occupational skills and to resolve psychological problems–such as substance abuse or aggression–that might interfere with their reintegration into society. . . . Since then, however, rehabilitation has taken a back seat to a 'get tough on crime' approach that sees punishment as prison's main function." Etienne Benson, *Rehabilitate or Punish?*, 34 AM. PSYCH. ASS'N 46, 46 (2003).

45. Ed Vulliamy, *Nixon's 'War on Drugs' Began 40 Years Ago, and the Battle is Still Raging*, GUARDIAN (July 23, 2011), https://www.theguardian.com/society/2011/jul/24/war-on-drugs-40-years#:~:text=Drugs-,Nixon's%20'war%20on%20drugs'%20began%2040%20years%20ago%2C%20and,the%20battle%20is%20still%20raging&text=Four%20decades%20ago%2C%20on,the%20%22war%20on%20drugs%22 [https://perma.cc/UN76-LNRE].

46. *Id.*

47. Comprehensive Drug Abuse Prevention and Control Act of 1970, 21 U.S.C. § 801 (1970).

48. DRUGWARFACTS.ORG, *Total Annual Arrests in the US by Offense Type in 2019 Compared With 1973*, https://drugwarfacts.org/node/233 (last visited Jan. 21, 2021) [https://perma.cc/Q6EN-3KTP].

49. PRISON POL'Y INITIATIVE, *Tracking State Prison Growth in 50 States* (2014), https://www.prisonpolicy.org/reports/overtime.html [https://perma.cc/8YPY-JY8R].

in assessing risk. Against this backdrop, we will consider what Bonta and Andrews (2007) describe as four generations of risk assessment[50] used in criminal justice: (1) professional judgment; (2) evidence-based tools; (3) evidence-based and dynamic assessments; and (4) systematic and comprehensive assessments. This Section will briefly survey these generations, and then discuss their applications.

A. Four Generations of Risk Assessment

The first generation of risk assessment is marked by the period up to and including the 1960s, in which assessment of risk was clinically rendered by professionals in correctional roles, as well as professionals trained in psychology and social work.[51] The goal was to predict "dangerousness," and first generation efforts were rife with false positives, with some studies in the 1960s and 1970s identifying between 54 and 99 percent of participants as "dangerous."[52] Because of its sole reliance on opinion, professional judgement was comparatively unstructured, and hence inconsistent, relative to the data-driven approaches that would come to dominate the risk assessment domain in the 1970s, commonly thought to be the beginning of the "tough on crime" era.[53]

The second generation of risk assessment was defined by evidence-based, or "actuarial," tools to consider individual characteristics for the purpose of assigning quantitative scores. As Skeem and Monahan (2011) write, "[n]o distinction in the history of risk assessment has been more influential than Paul Meehl's (1954) cleaving the field into 'clinical' and 'actuarial' (or statistical) approaches."[54] The ability to quantify recidivism risk based primarily on static factors such as criminal history was a step change in accuracy and consistency when compared to professional opinion. Bonta and Andrews (2007) point to two shortcomings.[55] First, this generation of tools relied on statistical correlations between recidivism and historical data, and did not predicate scores on establishment of causal linkage. Second, the static, backward-looking factors used to compute risk scores are immutable and assumed that only past behavior is considered for assessment of future risk. The Violence Risk Assessment Guide (VRAG)[56] and General Statistical Information for Recidivism (GSIR)[57] are examples of second-generation tools by virtue of their primarily static risk predictors.

---

50. James Bonta & D. A. Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation, 2006-2007*, PUB. SAFETY CAN. (2007).

51. *Id.*

52. JOHN MONAHAN, PREDICTING VIOLENT BEHAVIOR: AN ASSESSMENT OF CLINICAL TECHNIQUE 244, 246-50 (1981).

53. SENT'G PROJECT, *Criminal Justice Facts*, https://www.sentencingproject.org/criminal-justice-facts/ (last visited Jan. 21, 2021) [https://perma.cc/929X-72WM].

54. Jennifer L. Skeem & John Monahan, *Current Directions in Violence Risk Assessment* (Va. Pub. L. & Legal Theory, Research Paper Series No. 2011-13, 2013).

55. Bonta & Andrews, *supra* note 50.

56. Min Yang, Stephen C. P. Wong & Jeremy Coid, *The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools*, 136 PSYCH. BULL. 740, 742 (2010).

57. *Id.*

The third generation of risk assessment built upon research of the late 1970s and early 1980s. The new approach augmented scores based on historical, static data, by considering:

> [D]ynamic items investigating the offender's current and ever-changing situation. Questions were asked about present employment (after all, one can lose a job or find a job), criminal friends (one can make new friends and lose old friends), family relationships (supportive or unsupportive), etc. The third generation risk instruments were referred to as 'risk need' instruments and a few of these were also theoretically based.[58]

This era of tools was marked by consideration of "criminogenic needs," referring to factors statistically correlated to recidivism.[59] Criminogenic needs in this era were generally comprised of categories referred to as the "Big Four": antisocial behavior, antisocial personality pattern, antisocial associates, and antisocial cognitions.[60] This generation of risk instrument introduced the risk-need-responsivity model,[61] in which the risks and needs of the offender influence decisions about rehabilitative programs and the environments into which they are placed. The main goal of this evolved approach is to consider an offender's changing life circumstances and to understand dynamic risk factors. This understanding would lead to targeted interventions against specific risk factors, thus theoretically reducing an offender's risk profile. Examples of third generation tools are Level of Service Inventory-Revised (LSI-R),[62] Historical, Clinical, and Risk Management Violence Risk Assessment Scheme (HCR-20),[63] and the Violence Risk Scale (VRS).[64]

The mid-2000's saw the arrival of fourth generation risk assessment tools, which are more sophisticated based on their inclusion of a broader range of criminogenic needs/factors related to life circumstances that were previously not considered. The "Big Four" criminogenic needs categories evolved to become the "Central Eight": family dynamics, work/vocation, substance abuse, leisure/recreation, peer relationships, emotional stability/mental health, criminal attributes, and residential stability.[65] This level of sophistication enables more fine-grained measures of the probability that an offender will re-offend, but as we have observed in data science generally, increasing sophistication in predictive models leads to decreasing explainability. The application of advanced approaches, like machine learning, to enable models to dynamically adjust to new data has heightened the explainability challenge. Examples of fourth generation tools are Level of Service/Case Management

58. Bonta & Andrews, *supra* note 50.

59. Chris Baird, *Criminogenic Needs*, NAT'L COUNCIL ON CRIME & DELINQUENCY (Feb. 2017), https://www.nccdglobal.org/sites/default/files/criminogenic_needs.pdf [https://perma.cc/9FDA-4KJT].

60. Emma J. Palmer, Ruth M. Hatcher, James Mcguire & Clive R. Hollin, *Cognitive Skills Programs for Female Offenders in the Community: Effect on Reconviction*, 42 CRIM. JUST. & BEHAV. 345, 347 (2015).

61. Bonta & Andrews, *supra* note 50.

62. *See* JAMES BONTA & D. A. ANDREWS, THE LEVEL OF SERVICE INVENTORY-REVISED (1995).

63. Yang et al., *supra* note 56.

64. *Id.*

65. Baird, *supra* note 59.

Inventory (LS/CMI),[66] Correctional Offender Management Profiling for Alternative Sanctions (COMPAS),[67] and the federal Post Conviction Risk Assessment (PCRA).[68]

The primary application of assessment instruments has generally been in support of decisions regarding FTA risk and pretrial recidivism risk; but over the course of these generational shifts in risk assessment instruments, their application has expanded, with instruments used today for FTA risk, pretrial risk for both recidivism and violent recidivism, objective classification, sentencing, and parole. Most assessment instruments are driven by classification algorithms used to group offenders by risk levels of low, medium, and high. In 2019, the Electronic Privacy Information Center (EPIC) conducted a survey of state practices and usage,[69] and estimated that there are approximately 35 assessment instruments in use in the United States, spanning every state except Massachusetts and New Hampshire, where usage is currently the subject of evaluation and debate.[70] The instrument landscape is a mix of commercial off-the-shelf products adapted to local jurisdictions and custom-built instruments, tailor-made for a particular jurisdiction's use. As EPIC describes, "[t]he functions vary between pre-trial, sentencing, prison management, and parole. Most of these tools, including their existence, are largely opaque and change often."[71] The economics of risk assessment in the last decade have become increasingly favorable, with cheaper access to and storage of data and cheaper computation leading to more resources available for algorithmic decision support systems. The landscape of tools and their applications is highly variable across jurisdictions, with no consistent set of guidelines or statutory requirements beyond constitutional requirements for due process and equal protection.

B. Applications

1. Pretrial

The broadest application of risk assessment instruments is in the pretrial phase, for assessment of FTA risk, risk of recidivism, and risk of violent recidivism. As Foote (1956) famously said, "[p]retrial decisions determine mostly everything."[72] While fairness has always been a foundational goal of the criminal justice system, the

66. *See* JAMES BONTA, D. A. ANDREWS & J. STEPHEN WORMITH, THE LEVEL OF SERVICE/CASE MANAGEMENT INVENTORY (2004).

67. EQUIVANT, *Northpointe Suite Risk Need Assessments*, https://www.equivant.com/northpointe-risk-need-assessments/ (last visited Jan. 21, 2021) [https://perma.cc/QJH8-Z58V].

68. U.S. COURTS, *Post Conviction Risk Assessment*, https://www.uscourts.gov/services-forms/probation-and-pretrial-services/supervision/post-conviction-risk-assessment (last visited Jan. 21, 2021) [https://perma.cc/5R8B-5XZZ].

69. ELEC. PRIV. INFO. CTR., *Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools*, https://epic.org/algorithmic-transparency/crim-justice/ (last visited Jan. 21, 2021) [https://perma.cc/BN5S-PUEG].

70. *Id.*

71. *Id.*

72. Candace McCoy, *Caleb Was Right: Pretrial Decisions Determine Mostly Everything*, 12 BERKELEY J. CRIM. L. 135, 135 (2007).

continued pursuit of accuracy is underscored by the generational shifts in assessment approach, and increasingly, technology. To that end, use of algorithmic tools in criminal procedure has its origins in FTA assessment, which is simply a binary decision regarding whether or not an individual is likely to return for a scheduled court appearance: high-risk individuals for FTA are detained, while the remainder are released, either on their own recognizance or after posting bail. While all fifty states and the District of Columbia maintain constitutional and statutory guidelines for pretrial release eligibility,[73] judges exercise additive discretion in order to protect victims and the public, often with the help of assessment risk scores. Pretrial assessments expanded into recidivism risk estimates as tools and approaches increased in complexity and sophistication, with the intent of giving judges the ability to differentiate low-risk individuals from high-risk individuals, and to understand who among the latter was likely to be violent.

The risk assessment is generally administered via questionnaire following an arrest and booking, often without a lawyer present. In the case of COMPAS, a survey of 137 questions provides the inputs for an algorithmic model, the output of which will have long-lasting impact on the life of the defendant. It is not uncommon for an indigent defendant to meet their attorney at the bail hearing, at which neither are assured visibility into what resources the judge may or may not use to make a detention or bail decision. At that hearing, however, a classification of low risk, medium risk, or high risk will set in motion a cascading chain of consequential, and often irrevocable, decisions.

2. Sentencing

Sentencing is seeing an increasing use of assessment tools. As discussed above, the use of an assessment instrument for sentencing, judicial reliance on algorithmic tools, and the black box problem were the central points of contention in a landmark 2016 Wisconsin Supreme Court case, *State v. Loomis*.[74] Eric Loomis, a defendant arrested for operating a vehicle used in a drive-by shooting, pled guilty to a lesser charge and was assessed for pretrial recidivism risk, general recidivism risk, and violent recidivism risk, scoring high on all three. At sentencing, the instrument used by the judge, COMPAS, did not allow for transparency into the calculation of Loomis's high-risk scores. Loomis sued, alleging violation of his due process rights. The court disagreed and held that use of an assessment instrument in sentencing is permitted, as long as its documented purpose is decision support and not sole reliance.[75]

The case gave rise to debates about what degree of reliance was appropriate and the role of self-interest. Kehl, Guo, and Kessler (2017) write, "[the judge] may simply

73. NAT'L CONFERENCE STATE LEGISLATURES, *50 State Chart | Pretrial Release Eligibility* (Mar. 13, 2013), https://www.ncsl.org/research/civil-and-criminal-justice/pretrial-release-eligibility.aspx [https://perma.cc/5AWW-6WDF]; *see also* NAT'L CONFERENCE STATE LEGISLATURES, *Pretrial Detention* (Jun. 7, 2013), https://www.ncsl.org/research/civil-and-criminal-justice/pretrial-detention.aspx [https://perma.cc/66PS-RKUQ].

74. State v. Loomis, 881 N.W.2d 749, 752-53 (Wis. 2016).

75. *Id.* at 768.

take a risk-averse approach and impose more stringent sentences on criminals who are labeled high risk in order to avoid potential blame for a high-risk criminal who received a less severe sentence and ultimately did reoffend." In a jurisdiction like Wisconsin where judges are elected, they go on to note, "[i]t is difficult to imagine that a 'high risk' label will not result in a longer sentence."[76]

3. Classification

Following conviction and sentencing, more algorithms await offenders in the form of objective classification systems[77] to predict prisoner misconduct, assess the level of security required, and determine where a prisoner will be housed and the level of supervision they will receive. As Bench and Allen (2003) explain, "[a]lthough the influence of objective classification systems on the operation of the American prison system cannot be denied, it has received only moderate attention from researchers."[78]

Coincident with second generation risk assessment, the shift from subjective determination of institutional placement to an objective, formulaic approach took place in the 1980s. Since that time, all fifty states have implemented an objective classification system.

Risk assessment instruments are not used for objective classification. As Austin (2003) explains, risk assessment instruments:

> [H]ave been normed on samples of persons placed on probation or parole based on their arrest, supervision violation, or re-incarceration rate and should not to be used for making custody/security designations. Although some of the factors used in risk assessment are the same factors used for prison classification, there are several that either do not apply (e.g., current employment status, current marital status, etc.) or are not predictive of prison conduct (e.g., age at first arrest, associations with criminal peer groups, etc.).[79]

There are two systems within this domain, external and internal. Austin (2003) writes:

> External classification places a prisoner at a custody level that will determine where the prisoner will be housed. Once the prisoner arrives at a facility, internal classification determines which cell or housing unit, as well as, which facility programs (e.g., education, vocational, counseling, and work assignments) the prisoner will be assigned.[80]

The inputs to these algorithmic systems are driven not by self-administered surveys or questionnaires but by official documentation of age, gender, histories of

---

76. Danielle Kehl, Priscilla Guo & Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, RESPONSIVE COMMUNITIES INITIATIVE, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y, HARV. L. SCH. (2017).

77. James Austin, *Findings in Prison Classification and Risk Assessment*, FED. BUREAU PRISONS, U.S. DEP'T JUSTICE (2003).

78. Lawrence L. Bench & Terry D. Allen, *Investigating the Stigma of Prison Classification: An Experimental Design*, 83 PRISON J. 367, 368 (2003).

79. Austin, *supra* note 77.

80. *Id.*

violence and mental illness, gang membership, and recent (past 12 months) disciplinary actions. Periodic reassessment during incarceration is common.[81]

4. Reentry and Parole

Offenders eligible for parole are subject to a pretrial release assessment to advise on reentry and parole risk. According to the U.S. Courts Probation and Pretrial Services Office, "Federal judiciary policy in the 1970s required probation officers to 'classify persons under supervision into maximum, medium, and minimum supervision categories dependent upon the nature and seriousness of the original offense, extent of prior criminal history, and social and personal background factors in the individual case.'"[82] Today, this process is carried out using tools such as COMPAS, PCRA, and more situation-specific instruments such as Inmate Prerelease Assessment (IPASS),[83] for example, which is used for offenders in need of continued 12-step attendance and treatment entry. Some institutions specify and require certain risk instruments be used. As Conti-Cook (2019) explains, "COMPAS is embedded in the actual [New York State Corrections and Community Supervision] directive . . . they're intertwined. This is part of their regulation."[84] In any case, the approach is consistently actuarial: surveys and questionnaires provide inputs to algorithmic classifiers of risk and need.

It is worth noting that while sentencing and parole proceedings are equally represented in this discussion as steps in the criminal justice process, the procedural protections afforded offenders who find themselves in these two phases are not equivalent. As McGarraugh (2013) writes, "[d]espite the uncertainty regarding the precise limits of required procedural protection, several things are clear: the right to counsel, evidentiary standards, and the qualifications of the decision-maker are distinctly different at sentencing and parole."[85]

From the point of arrest, through conviction and incarceration, and to reentry into society, an offender could be the subject of as many as eight consequential decisions that are either determined or influenced by algorithms. FTA, pretrial recidivism, general recidivism, violent recidivism, sentencing, objective classification (both internal and external), risk of reentry, and parole supervision levels are all subject to actuarial calculations, in which the outputs of one algorithm can bias the inputs of a subsequent algorithm. It is a well-known phenomenon in data science that bias produces systematic errors, which are amplified and compounded in a sequence of

81. *Id.*

82. PROBATION & PRETRIAL SERVS. OFF., ADMIN. OFF. U.S. COURTS, *An Overview of the Federal Post Conviction Risk Assessment* (June 2018), https://www.uscourts.gov/sites/default/files/overview_of_the_post_conviction_risk_assessment_0.pdf [https://perma.cc/ZDF9-TMWP].

83. See David Farabee, Kevin Knight, Bryan R. Garner & Stacy Calhoun, *The Inmate Prerelease Assessment for Reentry Planning*, 34 CRIM. JUST. & BEHAVIOR 1188, 1188 (2007).

84. Cynthia Conti-Cook & Glenn Rodriguez, *FAT* 2019 Implications Tutorial: Parole Denied: One Man's Fight Against a COMPAS Risk Assessment*, YOUTUBE (Feb. 22, 2019), https://www.youtube.com/watch?v=UySPgihj70E (recording of panel at FAT* 2019 Conference).

85. Pari McGarraugh, *Up or Out: Why Sufficiently Reliable Statistical Risk Assessment Is Appropriate at Sentencing and Inappropriate at Parole*, 97 MINN. L. REV. 1079, 1084 (2013).

successive calculations. The errors produced by racial bias are especially troubling, as the dramatic increase in use of algorithmic tools in criminal justice is taking place against a decades-long backdrop of mass incarceration of African Americans.

III. THE BIAS PROBLEM

Bias in this context is a double entendre: it refers to the colloquial meaning of "a personal and sometimes unreasoned judgment,"[86] as well as the data science meaning of, "systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others."[87] When it comes to artificial intelligence's (AI)[88] potential risks and ethical challenges, bias in machine learning algorithms is among the discipline's most studied domains, and scholarly research in recent years reflects this expanded focus. While bias in machine learning was first introduced by Mitchell (1980),[89] the meaning has since become increasingly focused on biases relating to gender, race, and ethnicity reflected in training data.[90] Consequently, societal prejudices toward groups of people are reflected in data used to develop algorithmic models, which O'Neil describes as "weapons of math destruction."[91] The biases lurking within these models are impediments to fairness that appear in different forms, both in society and in the legal domain. Gender bias and stereotyping, for example, appear in our everyday language that is subsequently used to identify images in computer vision models that ascribe femininity to cooking.[92] These same stereotyping issues affect the AI discipline of natural language processing[93] applications used in personal assistants and conversational bots, which are likely to presume doctors are men and nurses are women.[94] The issue of racial bias in particular has proved problematic in the algorithmically-driven world of banking, with a recent Berkeley study showing that lenders charge otherwise-equivalent Latinx/African American

86. MERRIAM-WEBSTER DICTIONARY, *Bias*, https://www.merriam-webster.com/dictionary/bias (last visited Jan. 21, 2021) [https://perma.cc/5HXS-98UL].

87. *Id.*

88. U.S. NAT'L INST. SCI. & TECH., *Artificial Intelligence*, https://www.nist.gov/topics/artificial-intelligence (last visited Jan. 21, 2021) [https://perma.cc/Q3PH-M2ZJ].

89. Tom M. Mitchell, *The Need for Biases in Learning Generalizations* (Rutgers Comput. Sci. Tech. Rept. CBM-TR-117, Rutgers U., 1980).

90. *See* Kate Crawford, *The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford #NIPS2017*, YOUTUBE (Dec. 10, 2017), https://www.youtube.com/watch?v=fMym_BKWQzk (Keynote Speech at NIPS 2017 Conference).

91. O'NEIL, *supra* note 20.

92. Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez & Kai-Wei Chang, *Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints*, arXiv preprint arXiv:1707.09457 (2017).

93. "[A] field of artificial intelligence that enables computers to analyze and understand human language. It was formulated to build software that generates and comprehends natural languages so that a user can have natural conversations with his or her computer." Jake Frankenfield, *Natural Language Processing (NLP)*, INVESTOPEDIA (updated Aug. 31, 2020), https://www.investopedia.com/terms/n/natural-language-processing-nlp.asp [https://perma.cc/ALZ4-49VQ].

94. *See generally* Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama & Adam T. Kalai, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*, 2016 ADVANCES NEURAL INFO. PROCESSING SYS. 4349 (July 21, 2016).

borrowers higher rates for purchase and refinance mortgages, costing $765M yearly.[95] Online commerce saw a similar effect in 2016, when Amazon introduced an upgrade to its Prime service, Prime Free Same-Day Delivery, in 27 metropolitan areas in the United States. Shortly thereafter, a Bloomberg analysis revealed that several of the cities included predominantly Black ZIP codes that were 50 percent less likely to have the new service.[96] These examples of numbers-driven decisionmaking are all too common: with no overt intent to discriminate, a myopic focus on a narrowly defined outcome nonetheless reinforces inequality along racial and socioeconomic lines. As Eubanks (2017) describes, "[i]t is mere fantasy to think that a statistical model or a ranking algorithm will magically upend culture, policies, and the institutions built over centuries."[97]

In August 2014, Attorney General Eric Holder spoke at the annual meeting of the National Association of Criminal Defense Lawyers.[98] In a speech touting the introduction of data-driven reforms, he also warned about the potential for unintended consequences, remarking:

> Legislators have introduced the concept of "risk assessments" that seek to assign a probability to an individual's likelihood of committing future crimes and, based on those risk assessments, make sentencing determinations. Although these measures were crafted with the best of intentions, I am concerned that they may inadvertently undermine our efforts to ensure individualized and equal justice. By basing sentencing decisions on static factors and immutable characteristics – like the defendant's education level, socioeconomic background, or neighborhood – they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.[99]

It foreshadowed what would come two years later, in May 2016, when ProPublica's Angwin, Larson, Mattu, and Kirchner (2016) published the results of an analysis of COMPAS assessments from the Broward County Sheriff's Office in Florida.[100] The study built on existing research, data from the Florida Department of Corrections, and over two years of Freedom of Information Act (FOIA) requests.[101]

---

95. UC BERKELEY RSCH., Press Release, Mortgage Algorithms Perpetuate Racial Bias in Lending, Study Finds (Nov. 13, 2018), https://news.berkeley.edu/story_jump/mortgage-algorithms-perpetuate-racial-bias-in-lending-study-finds/ [https://perma.cc/9D2T-K6JN].

96. David Ingold & Spencer Soper, *Amazon Doesn't Consider the Race of Its Customers. Should It?*, BLOOMBERG (Apr. 21, 2016), https://www.bloomberg.com/graphics/2016-amazon-same-day/ [https://perma.cc/7DF8-E6H4].

97. VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 178 (2017).

98. Eric Holder, Attorney General of the United States, Keynote Address at the National Association of Criminal Defense Lawyers 57th Annual Meeting (Aug. 1, 2014), https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th [https://perma.cc/M8G9-JVUP].

99. *Id.*

100. Angwin et al., *supra* note 15, at ¶¶ 12–15.

101. Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (summarizing field of research) [https://perma.cc/6YTS-ZYCU].

A. ProPublica and COMPAS

ProPublica's analysis of COMPAS represented a quantified realization of the long-held concern that historical racial bias in the criminal justice system would become encoded in algorithmic tools in use today. The findings revealed racial inequalities in risk scores, reporting:

> [T]he algorithm made mistakes with Black and white defendants at roughly the same rate but in very different ways. The formula was particularly likely to falsely flag Black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than Black defendants.[102]

The study focused on "Risk of Recidivism" and "Risk of Violent Recidivism," and the core findings were unambiguous: "Black defendants were 77 percent more likely to be pegged as at higher risk of committing a future violent crime and 45 percent more likely to be predicted to commit a future crime of any kind."[103] To eliminate the possibility that the disparity was simply due to criminal history and the type of crime for which individuals were arrested, the analysis included a logistical regression test that "isolated the effect of race from criminal history and recidivism, as well as from defendants' age and gender."[104]

The company that developed COMPAS, Northpointe (now Equivant), published a formal rebuttal response arguing, among other things, that measures of accuracy of risk scores should be considered relative to "base rates of recidivism," and that it is "not proper to make an assessment of racial bias" without this consideration.[105] In this context, "base rate" refers to the respective rates at which Black and white defendants are arrested for new crimes. Because Black people in the United States are twice as likely to be arrested than white people,[106] the "base rates" argument suggests that one should simply expect Black defendants to receive a higher score, commensurate with the difference in base rates.

In other words, Northpointe asserted that "fairness" is achievable with "accuracy"[107] if one accounts for this historical pattern of over-policing communities of color to predict who is likely to be arrested in the future. In reference to socially disadvantaged groups, Hannah-Moffat (2011) writes:

---

102. Angwin et al., *supra* note 15.

103. *Id.*

104. *Id.*

105. William Dieterich, Christina Mendoza & Tim Brennan, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, NORTHPOINTE (Jul. 8, 2016), https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html [https://perma.cc/GWK2-5P3R].

106. FED. BUREAU INVESTIGATION, U.S. DEPT. JUSTICE, *2018 Crime in the United States, Arrests by Race and Ethnicity* (2018) [hereinafter FBI Crime Statistics 2018], https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/tables/table-43 (highlighting that out of 7.7 million total U.S. arrests in 2018, 2.1 million, or 27.2 percent, were Black or African American, more than double the percentage of African Americans among the U.S. population) [https://perma.cc/72Z5-56WW].

107. *See supra* note 16 on the definition of accuracy.

> These issues are important to the ethics of decision making because the base rate estimates for recidivism may actually be lower in the general offender population than what is predicted on risk assessment instruments. This may result in the possibility that a more severe penalty is administered on the basis of a risk assessment tool that inflates the actual risk posed by certain groups of offenders.[108]

The ProPublica examination of COMPAS and Northpointe's subsequent response motivated a number of scholarly research efforts in the wake of the analysis to investigate the findings and answer the question posed by ProPublica's Angwin and Larson (2016): "[s]ince Blacks are re-arrested more often than whites, is it possible to create a formula that is equally predictive for all races without disparities in who suffers the harm of incorrect predictions? Working separately and using different methodologies, four groups of scholars all reached the same conclusion. It's not."[109] A team of Stanford researchers, Corbett-Davies and Pierson (2016), wrote, "[i]f Northpointe's definition of fairness holds, and if the recidivism rate for Black defendants is higher than for whites, the imbalance ProPublica highlighted will always occur," and went on to describe ProPublica's findings as mathematically "inevitable."[110] Another group of researchers, Kleinberg, Mullainathan, and Raghavan (2016), concluded concurrent achievement of fairness and accuracy are elusive, writing "[w]e formalize three fairness conditions that lie at the heart of these debates, and we prove that except in highly constrained special cases, there is no method that can satisfy these three conditions simultaneously."[111] Subsequent research from Flores, Lowenkamp, and Bechtel (2017) was highly critical of ProPublica's methodology and conclusions, stating:

> Just as medicine uses actuaries to inform patient prognoses and the auto insurance industry uses actuaries to inform probabilities of risky driving behavior, the COMPAS is based on an actuary designed to inform the probability of recidivism across its three stated risk categories. To expect the COMPAS to do otherwise would be analogous to expecting an insurance agent to make absolute determinations of who will be involved in an accident and who won't. Actuaries just don't work that way.[112]

Cowgill and Tucker (2017) referenced the ProPublica study in a proposal for use of counterfactual explanation to measure bias and fairness in machine learning,

---

108. Kelly Hannah-Moffat, *Actuarial Sentencing: An "Unsettled" Proposition*, 30 JUST. Q. 270, 277-78 (2013).

109. Julia Angwin & Jeff Larson, *Bias in Criminal Risk Scores is Mathematically Inevitable, Researchers Say*, PROPUBLICA (Dec. 30, 2016), https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say [https://perma.cc/YFH4-2287].

110. Sam Corbett-Davies, Emma Pierson, Avi Feller & Sharad Goel, *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/ [https://perma.cc/A9ZH-MJCC].

111. Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, arXiv preprint arXiv:1609.05807 (2016).

112. Anthony W. Flores, Kristin Bechtel & Christopher T. Lowenkamp, *False Positives, False Negatives, And False Analyses: A Rejoinder To* Machine Bias: There's Software Used Across The Country To Predict Future Criminals. And It's Biased Against Blacks, 80 FED. PROBATION 38, 45 (2016).

writing: "[e]ven if COMPAS were racially biased, it may not have affected defendants' outcomes. If judges were already predisposed to sentence in a COMPAS-like way – e.g., they agreed independently with COMPAS – the tool would have no effect."[113] These research efforts argue that risk assessment instruments are still a step forward, albeit a marginal one, in fundamental fairness relative to the status quo.

B. Beyond ProPublica

The debate around the ProPublica study warrants broader examination into research that delves specifically into the question of racial bias in algorithmic assessment instruments and its potential impact on outcomes, inclusive of studies done both before and after 2016.

Two common elements are congruent with the ProPublica/Northpointe debate: the examination of fairness criteria and the inability to satisfy multiple conditions simultaneously. As Chouldechova (2017) points out, "[t]he differences in false positive rates and false negative rates cited as evidence of racial bias by [ProPublica] are a direct consequence of applying a [risk assessment instrument] that satisfies predictive parity to a population in which recidivism [base rates] differ across groups."[114] Huq (2019) investigates the disconnect between technical definitions of fairness and real-world implications, particularly as it relates to continued racial stratification in criminal justice, asserting: "[r]ather than asking about abstract definitions of fairness, a criminal justice algorithm should be evaluated in terms of its long-term, dynamic effects on racial stratification. The metric of nondiscrimination for an algorithmically assigned form of state coercion should focus on the net burden thereby placed on a racial minority."[115] Hao and Stray (2019) take a unique approach, creating an interactive online tool, based on COMPAS, for readers to assume the role of a data scientist and obtain firsthand experience with the statistical conundrum created by differing approaches to "fairness."[116]

Another common theme revolves around the inescapability of patterns of racial bias in historical data, and the lack of understanding amongst risk assessment practitioners. As Benjamin (2019) writes, "the practice of codifying existing social prejudices into a technical system is even harder to detect when the stated purpose of a particular technology is to override human prejudice."[117] In the context of sentencing decisions affected by actuarial risk instruments to predict general recidivism, Hannah-Moffat (2011) writes:

---

113. Bo Cowgill & Catherine Tucker, *Algorithmic Bias: A Counterfactual Perspective* 1 (Working Paper: NSF Trustworthy Algorithms, Dec. 2017).

114. Chouldechova, *supra* note 11.

115. Huq, *supra* note 8.

116. Karen Hao & Jonathan Stray, *Can You Make AI Fairer Than A Judge? Play Our Courtroom Algorithm Game*, MIT TECH. REV. (Oct. 17, 2019), https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/ [https://perma.cc/ZL8M-B7XY].

117. RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE 96 (2019).

> Because the tools classify and promote interventions based on categories of offender risk (i.e. low, medium, high), risk technologies tend to de-individualize punishments and can shift and reorient sentencing practices in unanticipated ways. Moreover, legal and correctional professionals who use risk information in decision-making are unlikely to have considered the documented limitations about the science of risk, and frequently have only a limited understanding of the actuarial technologies they are using.[118]

The study goes on to reference automation bias in this context, noting that "[r]isk scores impart a moral certainty and legitimacy into the classifications they produce, 'allowing people to accept them as normative obligations and therefore scripts for action.'"[119] But most importantly, the study questions the ability to evade historical patterns in data to create an unbiased instrument:

> The fact that actuarial risk assessments are typically created from the case files of subpopulations of incarcerated offenders raises concerns about the ability of any instrument to make an unbiased prediction of risk. Prison populations are not random; they are the products of past sentencing policies and patterns and they disproportionately represent Blacks, Aboriginals, and other socially disadvantaged groups.[120]

Proposals to mitigate biases in data are predicated on the existence of a net benefit to actuarial assessment generally, but Harcourt (2010) is not optimistic, stating that "risk today has collapsed into prior criminal history, and prior criminal history has become a proxy for race. The combination of these two trends means that using risk-assessment tools is going to significantly aggravate the unacceptable racial disparities in our criminal justice system."[121] Mayson (2019) goes further, questioning the usefulness of any attempt at prediction in this context, writing:

> All prediction looks to the past to make guesses about future events. In a racially stratified world, any method of prediction will project the inequalities of the past into the future. . . . Algorithmic risk assessment has revealed the inequality inherent in all prediction, forcing us to confront a problem much larger than the challenges of a new technology. Algorithms, in short, shed new light on an old problem.[122]

The law prohibits inclusion of race as a factor for consideration, but "[i]t's difficult to hide sensitive attributes from algorithms."[123] In reference to a COMPAS assessment, Benjamin (2019) notes that "the survey measures the extent to which an individual's life chances have been impacted by racism without ever asking an individual's race."[124] Race as a consideration is unwittingly encoded in data through

118. Hannah-Moffat, *supra* note 108.
119. *Id.* (quoting RICHARD V. ERICSON & KEVIN HAGGERTY, POLICING THE RISK SOCIETY (1997)).
120. *Id.* (omitting reference to cited studies).
121. Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237, 237 (2015).
122. Mayson, *supra* note 8, at 2218.
123. Author interview with Ece Kamar, Senior Researcher, Microsoft Research (January 2020).
124. BENJAMIN, *supra* note 117.

strong correlations between attributes like ZIP codes, income levels, credit ratings, and even a person's name,[125] among others. These immutable characteristics are often used as proxies for race and socioeconomic status, unbeknownst to the instrument's user because of the black-box nature of modern predictive tools. As these models become more sophisticated and increasingly reliant on more cheaply available data, stronger correlations to race will be an unintended consequence of getting "better." Furthermore, risk assessment algorithms are what statisticians call a "dynamic model," in which more data leads to more learning, resulting in more model tweaks and adjustments. The degree to which different commercial vendors incorporate learning in the form of model updates is unknown, but the complexity of factors influencing the act of re-offending are constantly in motion. Additionally, there is no "ground truth" to which one can compare predictions that lead to pretrial detention. If an offender was falsely judged to be "high risk" for FTA and consequently detained, there is no way to know whether or not such a decision represented a false positive.

IV. IMPACT ON CRIMINAL JUSTICE

In the context of criminal justice, biases have long influenced decisions about how and where to deploy law enforcement resources, with low-income and minority communities bearing the brunt of "over-policing," resulting in arrest statistics that are misaligned with general population demographics. African Americans comprise thirteen percent of the U.S. population,[126] but represent twenty-seven percent of the arrests.[127] Predictive policing tools encode these biases in algorithms to effectively automate decisions about where law enforcement conducts patrols and who will get arrested. While racial bias in predictive policing is not the subject of this Article, it does provide important context around who is more likely to be arrested, assessed for risk, and possibly detained.

A. Assessment, Detention, and Conviction

Studies correlating pretrial decisions to subsequent impacts are not new. Referring to the impact of bail on indigent defendants, Foote (1956) wrote, "[t]here is an extraordinary correlation between pretrial status (jail or bail) and the severity of the sentence after conviction, the jailed defendant being two or three times more likely to receive a prison sentence."[128] The cascading effects of algorithmic assessment and pretrial detention is quantified through synthesis of literature on specific elements of the criminal justice process. This process begins when an individual has been arrested

125. O'NEIL, *supra* note 20.

126. U.S. CENSUS BUREAU, *US Census Bureau Population Estimates* (July 1, 2018), https://www.census.gov/quickfacts/fact/table/US/PST045218 (Black or African American alone; includes persons reporting only one race) [https://perma.cc/A87F-SREM].

127. *See FBI Crime Statistics 2018*, *supra* note 106 (7.7 million total U.S. arrests in 2018, 2.1 million, or 27.2 percent, were Black or African American, more than double the percentage of African Americans among the U.S. population).

128. Caleb Foote, *The Coming Constitutional Crisis in Bail*, 113 U. PA. L. REV. 959, 960 (1965) (citing studies of the effects bail in Philadelphia, PA and New York, NY).

and charged, and the assessment survey is administered. It is followed by a hearing for those who can't post bail immediately, which is typically very brief—sometimes as little as thirty seconds[129]—to review charges, read rights, assign counsel, and make the bail/release/detention decision. In this hearing, the judge considers a number of factors including the seriousness of the charge, strength of the evidence, flight risk, and ability to post bail.[130] The "score" of the risk assessment for FTA and bail decisions is increasingly used as a decision support tool in U.S. courtrooms.

Given the presence of racial bias in pretrial risk assessment instruments, Black defendants have a greater likelihood of receiving a "high risk" classification, and hence pretrial detention, either because bail is denied or set a level that is financially unreachable. The New York state division of criminal justice did a 1995 review of disparities in processing felony arrests and found that in some parts of New York, Black defendants are thirty-three percent more likely to be detained awaiting felony trials than whites facing felony trials.[131] Black youth are also vulnerable to cascading effects of pretrial detention. As Sikmund (2004) reports: "[w]hile youth of color represent about a third of the youth population, . . . they represent 61 percent of detained youth. Youth of color are disproportionately detained at higher rates than whites, even when they engage in delinquent behavior at similar rates as white youth."[132] An exploration of cascading effects starts with the impact of pretrial detention on likelihood of conviction, and it begins with a defendant's circumstances. Most jurisdictions in the United States have rules to ensure constitutional and statutory protections in the form of a speedy trial; but even with these rules, any time at all behind bars has cascading effects. Because ninety-eight percent of cases are plea-bargained[133] and never go to trial, pretrial detention often begins with an asymmetrical negotiation between a prosecutor with the ability to use a defendant's time in detention as leverage and a defendant who has suddenly lost contact with family, employment, and other sources of social stability.

The subject of negotiation is the plea bargain, an offer of a reduced charge in exchange for a guilty plea, and many defendants take it. As Michelle Alexander explained in a 2017 documentary on Kalief Browder, "[a]lmost everybody pleads out, not because everybody is guilty, [but] because people crumble under the

129. Marie VanNostrand, *Pretrial Decisions Determine Mostly Everything*, LUMINOSITY, YOUTUBE (Apr. 5, 2014), https://www.youtube.com/watch?v=IIqFx7GE-HU.

130. Will Dobbie, Jacob Goldin & Crystal Yang, *The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges*, 108 AM. ECON. REV. 201, 206-09 (2018).

131. Bill Quigley, *Fourteen Examples of Racism in Criminal Justice System*, HUFFINGTON POST (Jul. 26, 2010), https://www.huffpost.com/entry/fourteen-examples-of-raci_b_658947 [https://perma.cc/G6FE-R4VC].

132. BARRY HOLMAN & JASON ZIEDENBERG, THE DANGERS OF DETENTION: THE IMPACT OF INCARCERATING YOUTH IN DETENTION AND OTHER SECURE FACILITIES, JUST. POL'Y INST. (2006), http://www.justicepolicy.org/images/upload/06-11_REP_DangersOfDetention_JJ.pdf (citing MELISSA SICKMUND, T.J. SLADKY & WEI KANG, CENSUS OF JUVENILES IN RESIDENTIAL PLACEMENT DATABOOK (2004)) [https://perma.cc/BWD6-N7QB].

133. *See* John Gramlich, *Only 2% of Federal Criminal Defendants go to Trial, and Most who do are Found Guilty*, PEW RSCH. CTR. (June 11, 2019), https://www.pewresearch.org/fact-tank/2019/06/11/only-2-of-federal-criminal-defendants-go-to-trial-and-most-who-do-are-found-guilty/ [https://perma.cc/U35B-UF8W].

pressure."[134] This leverage yields results: Williams (2003) finds that eighty percent of detained defendants are convicted, compared to sixty-six percent of released defendants.[135] Stevenson (2018) finds that pre-trial detention leads to a 6.2 % increase in the likelihood of conviction.[136] Dobbie, Goldin, and Yang (2016) find that pre-trial release decreases the probability of being found guilty by 15.6 % and decreases probability of pleading guilty by twelve percent.[137] While statistical relationships between pretrial detention, bail, and conviction are not in dispute, as Lum, Ma, and Baiocchi (2017) point out, "much of the literature in this area was only correlative and not causal,"[138] going on to "find a strong causal relationship between bail—an obstacle that prevents many from pre-trial release—and case outcome. Specifically, we find setting bail results in a thirty-four percent increase in the likelihood of conviction for the cases in our analysis."[139]

Regardless of whether a case results in a conviction, pretrial detention nonetheless has long-lasting effects relating to the interruption of employment, as contacts with the formal labor market are severed during incarceration, even if the incarceration is to await disposition of one's case. Dobbie et al. (2016) find that two years following the bail hearing, only 37.8 % of detained defendants are employed, versus 50.9 % of defendants who were released.[140] Conversely, pretrial release yields employment stability benefits, as a released defendant's probability of filing a tax return three to four years after the bail hearing increases by 4.3 %, and probability of employment in the same timeframe increases by 10.2 %.[141]

In the event of booking charges that do not result in a conviction, Lum, Boudin, and Price (2020) find that recommended levels of pretrial supervision (including detention) increased in twenty-seven percent of cases evaluated by a risk assessment instrument.[142] While the study was not intended to investigate racial differences in recommended supervision levels, it did reveal that Black defendants had a higher likelihood of the instrument's highest or second-highest risk category, writing:

> Disaggregating the analysis by race shows that while Black individuals received unwarranted charge-based exclusions and [new violent criminal activity] flags at a higher rate than non-Black individuals, they did not receive increased recommendations at a substantially higher rate due to the fact that Black individuals were more likely to be classified in the higher risk groups even before charge-based

134. TIME: THE KALIEF BROWDER STORY (Viacom 2017) (interviewing Michelle Alexander).

135. Marian R. Williams, *The Effect of Pretrial Detention on Imprisonment Decisions*, 28 CRIM. JUST. REV. 299, 303 (2003).

136. Megan T. Stevenson, *Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes*, 34 J.L. ECON. & ORG. 511, 532 (2018).

137. Dobbie et al., *supra* note 130.

138. Kristian Lum, Erwin Ma & Mike Baiocchi, *The Causal Impact of Bail on Case Outcomes for Indigent Defendants in New York City*, 3 OBSERVATIONAL STUD. 39, 39 (2017).

139. *Id.*

140. Dobbie et al., *supra* note 130.

141. *Id.*

142. Kristian Lum, Chesa Boudin & Megan Price, *The Impact of Overbooking on a Pre-Trial Risk Assessment Tool*, *in* PROCEEDINGS OF FAT*2020: THE ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (2020).

> increases are applied. This finding in and of itself may be worrisome to those who hope that risk assessment will close the racial gap in criminal justice outcomes.[143]

B. Sentencing and Classification

For detained defendants that are convicted, sentencing and classification determine both the length of incarceration and level of custody (e.g., medium security vs. maximum security). These decisions are algorithmically influenced at a minimum, but often decided by instruments, with the linkage to pretrial detention firmly established. Defendants who are detained pretrial are more likely to be sentenced to incarceration and have longer sentences than those who are released.[144] Dobbie et al. (2018) revealed a similar finding: compared to defendants who were released, defendants who were detained pretrial were fifteen percent more likely to be incarcerated, and received prison sentences that were 264.6 days longer on average.[145] Risk assessment instruments are routinely used in sentencing decisions, in concert with sentencing guidelines, and use immutable characteristics to arrive at a risk score. In Virginia, the Commonwealth developed an instrument specifically for sentencing, which "tallies demerits for past crimes with additional penalties for demographic characteristics found to be correlated with the commission of crime. Thus, a young, unemployed, never-married man is considerably more likely to face jail time than an older, divorced woman who held a job prior to committing an identical crime."[146]

Correlations to race are embedded in these characteristics and are consequently encoded in the calculations used to produce a score. Studies of sentence as a function of race have shown a statistical correlation. Using data compiled by the United States Sentencing Commission (USSC), Doerner and Demuth (2010) studied the joint effects of immutable characteristics, such as race, gender, and age on sentencing decisions in U.S. federal courts, and found that "Hispanics and Blacks, males, and younger defendants receive harsher sentences than whites, females, and older defendants after controlling for important legal and contextual factors,"[147] adding that "young Black male defendants receive the longest sentences."[148] The USSC stated in 2010 that in the federal system Black offenders receive sentences that are ten percent longer than white offenders for the same crimes.[149] The Sentencing Project reports that African Americans are twenty-one percent more likely to receive mandatory

143. *Id.*

144. Marian R. Williams, *The Effect of Pretrial Detention on Imprisonment Decisions*, 28 CRIM. JUST. REV. 299 (2003) (citing Charles E. Ares, Anne Rankin & Herbert Sturz, *The Manhattan Bail Project: An Interim Report on the Use of Pretrial Parole*, 38 N.Y.U. L. REV. 67 (1963)).

145. Dobbie et al., *supra* note 130.

146. Brian Netter, *Using Groups Statistics to Sentence Individual Criminals: An Ethical and Statistical Critique of the Virginia Risk Assessment Program*, 97 J. CRIM. L. & CRIMINOLOGY 699, 701 (2007).

147. Jill K. Doerner & Stephen Demuth, *The Independent and Joint Effects of Race/Ethnicity, Gender, and Age on Sentencing Outcomes in U.S. Federal Courts*, 27 JUST. Q. 1, 1 (2010).

148. *Id.* at 20.

149. Quigley, *supra* note 131.

minimum sentences than white defendants and twenty percent more likely to be sentenced to prison than white drug defendants.[150]

Objective classification algorithms ingest and magnify bias by using outputs of previous steps in the process—charge(s) for which the defendant was convicted and the resulting sentence—to influence the inputs for the subsequent step of determining custody level for incarceration and the facility programs for which the defendant will be placed. The custody level—minimum, medium, close, maximum—is determined by associations between prisoner attributes and prisoner misconduct. These attributes include "predictive factors," such as history of violence, mental illness, gang membership, non-participation in programs, and recent disciplinary actions.[151] Classification assessments also include factors that are "non-predictive" of misconduct, such as length of sentence and severity of offense.[152] The link between pretrial detention and these factors is easily inferred; any time behind bars subjects a defendant to risk of exposure to violence, and thus greater likelihood of documented disciplinary infractions while detained.

If defendants who are detained pretrial receive longer sentences on average, then the length of a sentence as an input to a classification algorithm cascades the bias from one domain (pretrial detention) to another (custody level for incarceration). A secondary classification assessment is conducted to advise on prisoner enrollment into facility programs, such as education programs, vocational training, counseling, and work assignments. This decision is also influenced by immutable factors in which racial disparity is encoded: education level, employment status, and history. Education and vocational training are desirable options for incarcerated persons who, for example, intend to use their time behind bars to prepare for successful reentry into society. However, a classification assessment that limits or directs an offender away from educational programs will indirectly increase the probability that they will re-offend, as Fasenfest and Case (2004) discovered, writing, "having a college education or vocational training decreased recidivism more than high school/GED training."[153]

C. Incarceration and Parole

There is a large body of literature spanning several decades on the multitude of social and psychological dynamics surrounding the incarceration experience, but this Article will focus on criminogenic factors that serve as inputs to assessment instruments. Toman et al. (2018) conducted a study to assess the effects of time spent in pretrial detention on both the likelihood and seriousness of prison misconduct, revealing "an association between pretrial detention length and inmate misconduct during time spent in state prison. As inmates serve longer terms in pretrial detention,

150. *Id.*

151. Austin, *supra* note 77.

152. *Id.*

153. David Fasenfest & Patricia Case, *Expectations for Opportunities Following Prison Education: A Discussion of Race and Gender*, 55 J. CORRECTIONAL EDUC. 24, 25 (2004).

their general likelihood and seriousness of offending increased."[154] Pretrial detention is an incarceration experience unto itself, and for defendants that are convicted, the effects of that experience are effectively cascaded to the domain of prison.

Once incarcerated, objective classification continues, with periodic "reclassification" assessments to possibly change a prisoner's level of custody. As Austin (2003) describes, "a reclassification form is used to score the prisoner on factors such as the type and number of misconduct reports lodged against the prisoner, the prisoner's participation in a variety of programs offered by the prison system, and the prisoner's work performance."[155] There exists a risk of self-fulfilling prophecy: exposure to violence is not something an offender can necessarily control, and involvement in what constitutes "misconduct" in classification assessment terms correlates to security level. Additionally, Bench and Allen (2003) report that "[t]he rate of disciplinary involvement may be directly influenced by offenders 'living up' to the expectations of a specific classification designation."[156] The study similarly revealed that "maximum-security offenders who stand out on a number of dimensions such as length of sentence, severity of offense, prior incarcerations, and propensity for violence can be housed in medium-security environments with no increased risk of disciplinary involvement."[157] In other words, many offenders are "overclassified."

A primary purpose of classification is to act on algorithmic predictions of likelihood to re-offend in the future, but Chen and Shapiro (2007) find that "[i]nmates housed in higher security levels are no less likely to recidivate than those housed in minimum security; if anything, our estimates suggest that harsher prison conditions lead to more post-release crime."[158] This finding was validated by Gaes and Camp (2009), who found:

> [I]nmates with a level III security classification who were randomly assigned to a security level III prison in the California prison system had a hazard rate of returning to prison that was 31 percent higher than that of their randomly selected counterparts who were assigned to a level I prison. Thus, the offenders' classification assignments at admission determined their likelihood of returning to prison.[159]

Incarceration includes exposure to drugs and substance abuse, and often begins well before entry into the criminal justice process. According to the Center for Prisoner Health and Human Rights, between sixty-three and eighty-three percent of arrestees had drugs in their system at the time of arrest, while Columbia University's

---

154. Elisa L. Toman, Joshua C. Cochran & John K. Cochran, *Jailhouse Blues? The Adverse Effects of Pretrial Detention for Prison Social Order*, 45 CRIM. JUST. & BEHAV. 316, 316 (2018).

155. Austin, *supra* note 77.

156. Lawrence L. Bench & Terry D. Allen, *Investigating the Stigma of Prison Classification: An Experimental Design*, 83 PRISON J. 367, 370-71 (2003).

157. *Id.*

158. M. Keith Chen & Jesse M. Shapiro, *Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach*, 9 AM. L. & ECON. REV. 1, 1 (2007).

159. Gerald G. Gaes & Scott D. Camp, *Unintended Consequences: Experimental Evidence for the Criminogenic Effect of Prison Security Level Placement on Post-Release Recidivism*, 5 J. EXPERIMENTAL CRIMINOLOGY 139, 139 (2009).

National Center on Addiction and Substance Abuse estimates that only eleven percent of incarcerated individuals in need of substance abuse treatment receive it in jail or prison.[160] The Center on Addiction (2010) reports that "[o]f the 2.3 million inmates crowding our nation's prisons and jails, eighty-five percent were substance-involved; 1.5 million met the DSM-IV medical criteria for substance abuse or addiction."[161] While not a result of a biased risk assessment, these conditions lay the groundwork for negatively impacting any future assessment, as "Substance Abuse" is a commonly referenced criminogenic factor in assessment instruments that potentially worsens with a longer period of incarceration if inmates are not receiving proper treatment.

Gangs are a source of regulation and social order in prison. For inmates who do not belong to a gang at the beginning of their sentence, incarceration increases the likelihood that they will join one. As Skarbek (2014) writes:

> Not only do many inmates feel they must join a gang, but gangs even issue written rules about appropriate social conduct. . . . In short, prison gangs form to provide extralegal governance. They enforce property rights and promote trade when formal governance mechanisms don't. They provide law for the outlaws.[162]

This is in part an artifact of mass incarceration: with the significant growth of the prison population in America in recent decades, inmates could no longer rely on unwritten tenets and social contracts that governed prison life to keep them safe.[163] Gang membership as a matter of perceived necessity while incarcerated can consequently have a detrimental effect on an inmate's likelihood of parole and the algorithmic assessment instrument that informs the decision.

Parole and reentry decisions are increasingly supported by algorithmic risk assessment tools, using factors which are consistent with pretrial assessments such as mental health, gang involvement, violent behaviors, negative social cognition, and substance abuse, among others, all of which are at risk of worsening simply by virtue of being incarcerated. Reentry instruments also take into account prison misconduct and disciplinary issues that occurred during incarceration.[164] The absence of a

160. Matt Gonzales, *Prisoners and Addiction*, DRUGREHAB.COM, ADVANCED RECOVERY SYSTEMS, https://www.drugrehab.com/addiction/prisoners/ (last visited Jan. 21, 2021) (summarizing studies of The Center for Prisoner Health and Human Rights and the National Center on Addiction and Substance Abuse at Columbia University) [https://perma.cc/455C-SZAF].

161. PARTNERSHIP TO END ADDICTION, *Substance Abuse and America's Prison Population 2010* (report can be downloaded at https://www.centeronaddiction.org/addiction-research/reports/behind-bars-ii-substance-abuse-and-america%E2%80%99s-prison-population) (last visited Jan. 21, 2021) (summarizing findings from its study under the organization's former name: CTR. ON ADDICTION, BEHIND BARS II: SUBSTANCE ABUSE AND AMERICA'S PRISON POPULATION (Feb. 2010)) [https://perma.cc/85D4-ST5J].

162. David Skarbek, *Why do prison gangs exist?*, OUP BLOG, OXFORD UNIV. PRESS (Aug. 6, 2014), https://blog.oup.com/2014/08/prison-gang-social-order/ [https://perma.cc/4A6N-D8PS].

163. J.D., *Why prisoners join gangs*, ECONOMIST (Nov. 12, 2014), https://www.economist.com/the-economist-explains/2014/11/12/why-prisoners-join-gangs [https://perma.cc/Y9E2-A6JV].

164. NORTHPOINTE, *Measurement & Treatment Implications of COMPAS Reentry Scales* (Mar. 30, 2009), https://www.michigan.gov/documents/corrections/Timothy_Brenne_Ph.D._Meaning_and_Treatment_Implications_of_COMPAS_Reentry_Scales_297503_7.pdf [https://perma.cc/VMT5-A9Y4].

feedback loop for correction of errors, and thus the absence of ground truth, is what causes algorithmic tools in this context to "blithely generate their own reality."[165] A biased pretrial risk assessment leads to a greater likelihood of conviction, leading to a greater likelihood of a longer sentence, leading to a greater likelihood of a higher level of custody, leading to greater risk of exposure to gangs, violence, and drugs, as well as greater risk of strains on mental health, all of which feed into the algorithmic instrument used in the parole process to assess an inmate's risk and needs. This is recursion.

The inmate's risk at the parole stage is calculated by an assessment instrument based on a set of factors that were potentially worsened during incarceration because of the output of the same instrument prior to incarceration. The presence of racial bias in the initial assessment creates a systematic error that propagates through the system, producing higher likelihoods of adverse impacts on African Americans on data ingested by multiple assessment instruments through the criminal justice process, including parole and into reentry.

D. Reentry

A formerly incarcerated person re-entering society has a set of widely researched challenges that closely correspond to the "Central Eight" criminogenic factors that define survey inputs to many modern, generation four risk algorithmic assessment instruments: family dynamics, work/vocation, substance abuse, leisure/recreation, peer relationships, emotional stability/mental health, criminal attributes, and residential stability.[166] Not only are formerly incarcerated individuals at greater risk of re-arrest and re-offending, the immutable factors that define their life circumstances will all but guarantee that any future risk assessment will continue to propagate the compounding of algorithmic bias that statistically contributed to their original arrest, detainment, conviction, sentence, and parole conditions.

The impact of incarceration is felt in the family. The majority of research into prisoners' families indicate a "heterosexual, nuclear family unit, usually consisting of an incarcerated father, a non-incarcerated mother, and young children."[167] One in four Black children experiences parental incarceration, which has negative impacts their health and education.[168] But most importantly, as Wakefield and Wildeman (2011) explain, "the long-term consequences of mass imprisonment might be for the intergenerational transmission of racial inequality."[169] This inequality begins with disparate policing, is reinforced algorithmically throughout the criminal justice process, and leaves its mark on the families of formerly incarcerated individuals. Incarceration

165. O'NEIL, *supra* note 20.

166. Baird, *supra* note 59.

167. Kolina J. Delgado, *The Impact of Incarceration on Families: A Summary of the Literature* 3 (Summer 2011), https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=1004&context=psych_student (omitting reference to studies) [https://perma.cc/CXH5-YDQD].

168. Huq, *supra* note 8, at 1110.

169. Sara Wakefield & Christopher Wildeman, *Mass Imprisonment and Racial Disparities in Childhood Behavioral Problems*, 10 CRIMINOLOGY & PUB. POL'Y 793, 795 (2011).

also places marital relationships under significant stress.[170] Mueller-Smith (2014) finds that "defendants serving three or more years in prison are at a statistically significant elevated risk of divorce post-release."[171] Siennick and Stewart (2014) report that, for marriages that occur before (versus during) incarceration, "each year of incarceration increases the odds that the inmate's marriage will end in divorce (before or after the inmate gets out of prison) by an average of thirty-two percent."[172] These circumstances will create a negative, reinforcing impact on any future assessment.

Employment is another persistent challenge for formerly incarcerated people. Mueller-Smith (2014) finds:

> [E]ach additional year of incarceration reduces post-release employment by 3.6 percent points. Among felony defendants with stable pre-charge income incarcerated for one or more years, reemployment drops by at least 24 percent in the five years after being released. Misdemeanor defendants show a small increase in take-up of cash welfare payments, and felony defendants show increases in Food Stamps benefits, which provide further evidence of lasting economic hardship post-release.[173]

This hardship often ensues despite participation in educational or vocational programs during incarceration. A study by Fasenfest & Case (2014) examined the usefulness of prison education in finding and maintaining employment after reentry, finding groups were divided by race:

> White males were more likely to perceive college courses in prison as being beneficial, reported a higher level of self-esteem post education, more often reported that they had taken courses post release to continue their education and were not likely to perceive barriers to employment post release. Black males reported opposite experiences that are likely reinforced by institutionalized racism that additionally reduces opportunities. Black males reported more value in vocational training that provided a work skill, experienced lower levels of post education self-esteem and reported more barriers to finding and maintaining employment.[174]

Simply finding employment is an obvious challenge for formerly incarcerated job applicants, starting with an application process that shows racial disparity. Pager (2005) finds that seventeen percent of white job applicants with criminal records received callbacks from employers while only five percent of Black job applicants with criminal records received callbacks.[175] Another cascading effect is disparity in

170. Delgado, *supra* note 167, at 4.

171. Michael Mueller-Smith, *The Criminal and Labor Market Impacts of Incarceration* 49-50 (Working Paper 18, 2018), http://www.columbia.edu/~mgm2146/incar.pdf [https://perma.cc/PMQ2-Q6F2].

172. COLLEGE CRIMINOLOGY & CRIM. JUST., FLA. STATE UNIV., Press release, FSU Criminologists Determine Why Prison Terms Make Couples More Likely to Divorce (May 30, 2014), http://criminology.fsu.edu/news/fsu-criminologists-determine-every-year-of-a-prison-term-makes-a-couple-32-percent-more-likely-to-divorce/ [https://perma.cc/3J9S-GXKK].

173. Mueller-Smith, *supra* note 171, at 4.

174. Case & Fasenfest, *supra* note 153, at 24.

175. Devah Pager, *Double Jeopardy: Race, Crime, and Getting a Job*, 2005 WIS. L. REV. 617, 642 (2005).

wage trajectory between Black and white workers once employment is obtained. Lyons and Pettit (2011) discovered that following reentry:

> The wages of Black ex-inmates grow about 21 percent more slowly each quarter after release than the wages of white ex-inmates. As we find no similar divergence in quarters leading up to incarceration, these results suggest that something about the experience of incarceration changes the relative wage trajectories of Blacks and whites.[176]

Substance abuse is a common factor for consideration in recidivism risk assessments. According to a 2007 study published in *The New England Journal of Medicine*, former inmates' risk of a fatal drug overdose is 129 times as high as it is for the general population during the two weeks after release.[177] A study by the Massachusetts Department of Public Health found that compared to the rest of the adult population, the opioid-related overdose death rate is 120 times higher for persons released from Massachusetts prisons and jails, and that nearly one of every 11 opioid-related overdose deaths were persons with histories of incarceration in Massachusetts jails and prisons. In 2015, nearly fifty percent of all deaths among those released from incarceration in Massachusetts were opioid-related.[178] While instance of substance abuse is not directly attributable to race, Black inmates have a higher risk of abuse without proper treatment due to longer periods of incarceration, compared to white offenders. Longer sentences are statistically correlated to pretrial detention and sentencing outcomes, both of which are decisions influenced by algorithmic instruments.

Being incarcerated has negative impacts on one's peer group and social network, due in no small part to the fact that incarcerated people form friendships in prison. Ouss (2011) has shown how peer effects from interactions while in prison result in learned patterns of criminality that influence criminal activity following reentry.[179] Bayer (2009) finds that "[t]he influence of peers primarily affects individuals who already have some experience in a particular crime category."[180] In studying indirect effects of the 2006 Italian prison pardon, Drago and Galbiati (2012) find that former inmates motivate criminal behavior within their peer group following reentry.[181] One's peer relationships and social network loom large as criminogenic factors used for assessment of risk.

---

176. Christopher J. Lyons, Becky Pettit, *Compounded Disadvantage: Race, Incarceration, And Wage Growth*, 58 SOC. PROBS. 257, 271 (2011).

177. Ingrid A. Binswanger et al., *Release from Prison — A High Risk of Death for Former Inmates*, 356 NEW ENGLAND J. MED. 157, 161 (Jan. 11, 2007).

178. MASS. DEP'T PUB. HEALTH, Data Brief, *An Assessment of Opioid-Related Overdoses in Massachusetts 2011-2015* 5 (Aug. 2017), https://www.mass.gov/files/documents/2017/08/31/data-brief-chapter-55-aug-2017.pdf [https://perma.cc/55FZ-7H3Y].

179. Aurelie Ouss, Prison as a School of Crime: Evidence from Cell-Level Interactions (Dec. 2011) (available at SSRN: https://ssrn.com/abstract=1989803 or http://dx.doi.org/10.2139/ssrn.1989803) [https://perma.cc/7ZNF-6EAR].

180. Patrick Bayer, Randi Hjalmarsson & David Pozen, *Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections*, 124 Q.J. ECONOMICS 105, 106 (Feb. 2009).

181. Francesco Drago & Roberto Galbiati, *Indirect Effects of a Policy Altering Criminal Behavior: Evidence from the Italian Prison Experiment*, 4 AM. ECON. J. 199, 200-201 (2012).

Residential stability is another key concern for formerly incarcerated people, as "release from jail or prison leaves a person particularly vulnerable to an episode of homelessness."[182] Metraux and Culhane (2004) found that rates of homeless shelter use are comparable among people exiting prison and people exiting state psychiatric hospitals.[183] Operators of homeless shelters commonly see recently incarcerated people among their residents, with one study reporting seventy percent of people in homeless shelters are formerly incarcerated, while another study reported fifty-four percent.[184] In Los Angeles and San Francisco, thirty to fifty percent of people released on parole and under supervision are homeless.[185] Like pretrial detention, long term incarceration represents a loss of contact with one's community, which typically serves as a source of social stability. The challenge is worsened by communities where "persistent poverty and lack of jobs and affordable housing make finding a permanent home difficult."[186]

Reentry into society following incarceration has many more challenges beyond the brief summaries provided in this Article, including effects relating to mental health, anger management, and what risk instruments refer to as "Criminal Personality," "Criminal Thinking," and "Non-compliance."[187] While one of the primary goals of the "Risk-Needs" approach to assessment is to identify where to provide interventions in pursuit of better outcomes, there is a self-reinforcing feedback loop that uses negative impacts on these life factors as inputs to whatever risk assessment occurs in the future. As Hannah-Moffat (2011) writes:

> Marginalized individuals' lives tend to be mired by a range of criminogenic and other needs, and consequently risk scores reflect systemic factors. High risk scores are associated with custodial sentences and/or a greater number of conditions attached to their disposition, making them more vulnerable to breach, increased surveillance, and further criminalization.[188]

E. Re-Arrested, Re-Assessed

In addition to having the world's largest prison population,[189] the United States has one of the world's highest recidivism rates. According to a recent study by the

182. STEPHEN METRAUX, CATERINA G. ROMAN & RICHARD S. CHO, INCARCERATION AND HOMELESSNESS, OFF. POL'Y DEV. & RSCH., DEP'T HOUSING & URBAN DEV. (2007), https://www.huduser.gov/portal/publications/homeless/p9.html [https://perma.cc/3WCP-8J7L].

183. Stephen Metraux & Dennis P. Culhane, *Homeless Shelter Use and Reincarceration Following Prison Release*, 3 CRIMINOLOGY & PUB. POL'Y 139, 150-51 (2004).

184. METRAUX ET AL., *supra* note 182, at 3.

185. JEREMY TRAVIS, AMY L. SOLOMON & MICHELLE WAUL, FROM PRISON TO HOME: THE DIMENSIONS AND CONSEQUENCES OF PRISONER REENTRY, URB. INST. 36 (June 2001), http://research.urban.org/UploadedPDF/from_prison_to_home.pdf [https://perma.cc/MX7C-PRR3].

186. *Id.*

187. NORTHPOINTE, *Practitioners Guide to COMPAS* (2012), http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf [https://perma.cc/H7ZR-H2TV].

188. Hannah-Moffat, *supra* note 108, at 286.

189. Roy Walmsley, *World Prison Population List, Twelfth Edition*, WORLD PRISON BRIEF 2 (2018), https://www.prisonstudies.org/sites/default/files/resources/downloads/wppl_12.pdf [https://perma.cc/6DH2-LS3G].

Bureau of Justice Statistics, "[f]ive in six (83 percent) state prisoners released in 2005 across 30 states were arrested at least once during the 9 years following their release."[190] Offenders who are re-arrested do not necessarily commit the same crime(s) that led to their initial incarceration. Mueller-Smith (2014) finds that "former inmates are especially likely to commit more property (e.g. theft or burglary) and drug-related crimes after being released, even if these crimes were not their original offenses,"[191] going on to write:

> [E]ach additional year that a felony defendant was incarcerated increases the probability of facing new charges post-release by 5.6 percentage points per quarter. What is particularly concerning about these results is that the incapacitation effect is disproportionately driven by misdemeanor charges, while the post-release criminal behavior shows mainly increases in felony offenses.[192]

A new arrest leads to a new risk assessment. In the case of COMPAS, the arrest is followed by a new 137-question survey[193] with questions about current charges and criminal history, followed by questions about life factors such as family, employment, and residential stability, all of which were likely negatively impacted by one's initial entry into the criminal justice system. The entire process begins again, and for Black defendants, it entails the same algorithmic bias compounded by new and incrementally more damaging answers to the assessment instrument's questions, leading to additional bias, which leads to additional compounding errors. There is no known process or means to control for or "de-bias" any of this history—it is permanent. For Black defendants, exposure to risk of injustice at the hands of algorithmic tools is, therefore, irreversible.

In a sense, risk assessment instruments assume every subject submits to the survey once; but this does not reflect the real world. The effects of a person's initial journey through the criminal justice system, including the challenges experienced in reentry such as unemployment or homelessness, are cascaded back into the system following re-arrest, through pretrial risk assessment, conviction, and sentencing. Commenting on use of risk assessment instruments in sentencing, Napa County Superior Court Judge Mark Boessenecker explains, "[a] guy who has molested a small child every day for a year could still come out as a low risk because he probably has a job, [but] a drunk guy will look high risk because he's homeless."[194] As it stands today, the cycle of data bias leading to compounding, systematic errors seemingly has no end; and once a defendant is in this cycle of injustice, there is no way out. In the era of big data and AI, this pattern of decisionmaking will become increasingly programmed in computer code, making it more efficient than ever before.

---

190. Mariel Alper, Matthew R. Durose & Joshua Markman, *2018 Update On Prisoner Recidivism: A 9-Year Follow-Up Period (2005-2014)*, BUREAU JUST. STAT. (May 23, 2018), https://www.bjs.gov/index.cfm?ty=pbdetail&iid=6266 [https://perma.cc/G7FN-PDA2].

191. Mueller-Smith, *supra* note 171, at 3-4.

192. *Id.* at 3.

193. *COMPAS Survey*, *supra* note 37.

194. Angwin et al., *supra* note 15, at ¶ 41.

V. FUTURE DIRECTIONS: RESEARCH AND POLICY

A broader analytical study is needed to quantify the impact of algorithmic compounding errors throughout the criminal justice process. A challenge remains due to the difficulty in establishing "ground truth," as Netter explains: "[t]he peril is that false positives are empirically incalculable because potential false positives are in jail."[195] That aside, an increasing reliance on statistical probabilities and data-driven risk instruments carries with it increasing access to data for researchers to understand cascading effects in this domain. The complexity of interactions between attributes is not quantitatively understood, and it may be unachievable without further study.

A. Towards a Research Agenda

Despite the large body of research on fairness and bias in predictive analytics and machine learning, there is very little in literature that examines the effect of errors due to bias that propagate through a sequence of algorithmic decision systems. Literature, with bespoke analyses of statistical correlations, references disparate impact as a consistent theme in studies of how algorithms continue to influence criminal justice, but with little elaboration on complex interactions and causal linkages between factors that determine a defendant's outcomes. A significant underlying problem is laid bare by Jacobs and Wallach (2019), who cite dubious claims by algorithmic decision systems to accurately measure "*unobservable theoretical constructs*—i.e., abstractions that describe phenomena of theoretical interest" which would include categories such as "creditworthiness," "teacher quality," or "risk to society."[196] Risk of recidivism and risk of prison misconduct are among the theoretical constructs inferred by instruments in the criminal justice process because of their relationship to observable properties like criminal history and the number of fights while incarcerated. In relation to COMPAS,[197] claims of reliance upon "actuarial science"[198] purport to address this disconnect, but the objectivity implied by this claim "obscure[s] the organizational, political, social, and cultural values . . . implemented in this system, masking inequality with a label of objectiveness."[199] In the criminal justice domain, this problem dates back to the second generation assessment era of the 1960's and 1970's, in which static factors like criminal history were the primary inputs into actuarial tools; but in today's fourth generation tools, the "Central Eight" criminogenic factors[200] play a substantial role in determination of a "measurement," such as a recidivism risk score. These factors are in large part immutable life circumstances that are reflected in the answers to assessment survey questions. Setting aside critiques of bias in the questions themselves, a single experience with the criminal

195. Netter, *supra* note 146, at 712.

196. Abigail Z. Jacobs & Hanna Wallach, *Measurement and Fairness*, arXiv:1912.05511 [cs.CY] 10 (2019) (emphasis in original).

197. EQUIVANT, *supra* note 67.

198. NORTHPOINTE, PRACTITIONERS GUIDE TO COMPAS CORE (Mar. 19, 2015), http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf [https://perma.cc/YFF9-6NJR].

199. Jacobs & Wallach, *supra* note 196, at 13.

200. Baird, *supra* note 59, at 1.

justice system makes future answers to survey questions more damaging to a defendant. Any subsequent experience with the system ingests any bias and/or unfairness from the previous experience and cascades it again.

Interrupting this cascade begins with a deeper understanding of causal interactions between assessment survey questions, subsequent bias, and the life circumstances that are impacted following reentry into society. Subsequent research requires a deeper exploration of these interactions, inclusive of causal links, social factors, the entrenched history of race in America, and how automation of consequential decisions is reshaping the criminal justice process.

Below is a proposal for three research questions to advance the understanding of cascading effects introduced in this Article:

1. What is the causal link between errors due to racial bias in early phases of criminal procedure and racially disparate outcomes in later stages and beyond?
2. How are bias errors amplified and compounded as they propagate through the criminal justice process?
3. Empirically speaking, how much do judges and magistrates rely on algorithmic instruments for consequential decisions?

1. Causal Linkage

While synthesis of literature reveals an inferred chain of statistical correlations, there is insufficient proof of causation; however, recent studies prove the feasibility of using modern causal inference methods to establish links between factors. In a study focused on the impact of bail on indigent defendants in New York City, Lum, Ma, and Baiocchi (2017) showed a causal link between setting bail and case outcome,[201] citing a number of other studies supported by modern methods for causal inference.[202] Heaton et al. (2017) studied misdemeanor cases in Harris County, TX, to establish a causal relationship between pretrial detention and both case outcomes and future crime, noting that "[o]ne key question for pretrial law and policy is whether detention actually causes the adverse outcomes with which it is linked, independently of other factors. On this question, past empirical work is inconclusive."[203]

---

201. Lum et al., *supra* note 138, at 1.

202. *Id.* Specifically, Lum et al. discussed the following studies:

> More recent work has applied modern methods for causal inference to assess the role played by the money bail system in the ultimate disposition of the case. Gupta et al. (2016) employ a measure of judge strictness as an instrumental variable to estimate the causal impact of setting bail on case outcome to data from Philadelphia and Pittsburgh. They find that assigning money bail increases the likelihood that the defendant is found guilty by 12%. Leslie and Pope (2016) applies a similar two-stage instrumental variable method to national-level data and data from New York City to assess the causal impact of pre-trial detention on case disposition. Leslie and Pope (2016) also uncovered a 'strong causal relationship.' Stevenson (2016) and Dobbie et al. (2016) employ similar methodology to data from Pennsylvania and Miami-Dade County. Both find a statistically significant impact of the money bail system on the outcome of the case.

Id. at 39.

203. Paul Heaton, Sandra Mayson & Megan Stevenson, *The Downstream Consequences of Misdemeanor Pretrial Detention*, 69 STAN. L. REV. 711, 714 (2017).

With this research problem as context, the study goes on to present evidence that does indeed use modern methods to quantify the causal effect of pretrial detention, finding that "defendants who are detained on a misdemeanor charge are much more likely than similarly situated releasees to plead guilty and serve jail time."[204]

While purposefully focused on narrow parameters, these studies demonstrate the potential for more rigorous quantitative methods, including expansion into a better understanding of the causal impact of racial bias in assessment instruments on the "Central Eight" criminogenic factors[205] commonly used in assessment surveys. As part of their Harris County study, Heaton et al. (2017) found:

> Although detention reduces defendants' criminal activity in the short term through incapacitation, by eighteen months post-hearing, detention is associated with a 30% increase in new felony charges and a 20% increase in new misdemeanor charges, a finding consistent with other research suggesting that even short-term detention has criminogenic effects.[206]

Debates between "accuracy" and "fairness" in the context of risk assessment instruments are likely to continue indefinitely, but mounting statistical evidence of racial bias in algorithmic systems provides an opportunity to move beyond statistical correlations and anecdotal evidence and into reliably proving causal connections between algorithmic instruments, near-term outcomes, and downstream cascading effects.

2. Amplification and Compounding

Few algorithms in a decisionmaking context exist in isolation or are impervious to ingesting or transporting bias to and from other algorithmic systems. Regardless, there is limited literature on the degree to which biases amplified by algorithms are then compounded as they make their way through a sequence of algorithmic systems. In this context, *bias amplification* refers to a phenomenon in which algorithmic models learn to overpredict negative outcomes and attributes for certain groups of people based on stereotypes, prejudices, and immutable factors like socioeconomic circumstances.[207] There is an existing body of work on bias amplification in machine learning, including research by Bolukbasi et al. (2016)[208] on gender stereotypes in natural language, and studies by Zhao et al. (2017),[209] Stock & Cisse (2017),[210] and Hendricks et al. (2018)[211] on bias in visual recognition systems. While computer

204. *Id.* at 717.

205. Baird, *supra* note 59.

206. Heaton et al., *supra* note 203, at 718.

207. Klas Leino, Matt Fredrikson, Emily Black, Shayak Sen, & Anupam Datta, *Feature-Wise Bias Amplification*, CARNEGIE MELLON UNIV. (2019), https://www.cs.cmu.edu/~mfredrik/papers/leino_iclr19.pdf [https://perma.cc/U7GL-K69M].

208. Bolukbasi et al., *supra* note 94.

209. Zhao et al., *supra* note 92.

210. *See generally* Pierre Stock & Moustapha Cisse, *ConvNets and ImageNet Beyond Accuracy: Understanding Mistakes and Uncovering Biases*, ECCV 2018 COMPUT. VISION 498 (2018).

211. *See generally* Lisa Anne Hendricks et al., *Women Also Snowboard: Overcoming Bias in Captioning Models*, ECCV 2018 COMPUT. VISION 793 (2018) (online version of chapter available at https://doi.org/10.1007/978-3-030-01219-9_47).

vision models are obviously not a focus of this Article, the mechanical underpinnings that lead them to amplify bias are precisely the same: algorithms capture, exploit, and magnify prejudices and stereotypes found in the data used to train them.

The transference and compounding of bias from one system to another in a decisionmaking sequence is less understood and represents a natural extension of existing bias amplification research. While this Article has relied upon proxy domains and lessons from complex systems to draw inferences, a controlled study and/or a simulation is required to isolate the effects of bias in each algorithmic instrument employed in key elements of the criminal justice process and to quantify their respective impact on subsequent elements in the process. Given the strong inference of causation between pretrial detention, for example, with criminogenic factors impacted post-release, such a study should extend beyond parole to capture the time horizon across which the full extent of harm is realized.

3. Judicial Reliance

The degree to which judges and bail magistrates rely on algorithmic instruments in their decisionmaking represents a gap in research. There is a body of work indicating that judges are less reliant on algorithmic instruments than the tools' critics suggest. In an ethnographic study focused on the reception of predictive algorithms among law enforcement and legal professionals, Brayne and Christin (2020) discovered "resentment toward predictive algorithms is fueled by fears of deskilling and heightened managerial surveillance."[212] The study found that in general, judges view algorithmic instruments much like they did mandatory sentencing in the 1980's: as a constraint on discretion. The study, however, also uncovered a displacement effect in which discretion is moved to less accountable areas of the criminal justice process, noting that "[n]ew actors also come into play, including data analysts, data entry specialists, and technology teams, who create novel forms of discretionary power within the institutions."[213]

Existing research on disparate impacts of pretrial detention periodically delve into the psychological forces at play in consequential decisionmaking. Consistent with findings from Kehl et al. (2017)[214] regarding risk aversion in sentencing decisions, Heaton et al. (2017) found that "individual judges or magistrates who make pretrial custody decisions suffer political blowback if they release people (either directly or via affordable bail) who subsequently commit violent crimes, but they suffer few consequences, if any, for setting unaffordable bail that keeps misdemeanor defendants detained."[215] Given the public pressures on judges and bail magistrates, it stands to reason that predictive tools would be a welcome aid to support high bail amounts and pretrial detention decisions in favor of public safety, but this is conjecture.

---

212. Sarah Brayne & Angèle Christin, *Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts*, SOC. PROBLEMS, OXFORD UNIV. 1 (2020).

213. *Id.*

214. Kehl et al., *supra* note 76.

215. Heaton et al., *supra* note 203, at 716-17.

Further research is needed to better understand the opposing forces represented by these two bodies of work. What is clear, however, is that there is no uniform application of algorithmic tools across jurisdictions, nor is there a consistent set of statutory guidelines that describe how judges should use these tools.[216]

B. Policy Interventions

In the meantime, there are a set of concrete policy positions immediately available to lawmakers that provide interventions, many of which can interrupt the interactions between decision points with fairness checks and curtail an escalating pattern of compounded racial injustice. First, transparency into increasingly opaque decision support systems must be required and ensured for all parties involved in any legal proceeding. Second, evidence rules must be revisited to address the exceptions that allow for machine testimony to be used in virtually every major element of criminal procedure other than the trial itself. Third, strict training requirements must be implemented for practitioners and users of these instruments, as recommended training from tool vendors is insufficient to ensure accountability and due process for defendants.

1. Opening the Black Box

The increasing use of data, statistics, and algorithms in criminal justice has surfaced a number of challenges not uncommon when introducing digital tools to analog-era legal processes and traditions. As Roth states, "[j]ust as human sources potentially suffer the so-called 'hearsay dangers' of insincerity, ambiguity, memory loss, and misperception, machine sources potentially suffer 'black box' dangers that could lead a factfinder to draw the wrong inference from information conveyed by a machine source."[217] But compared to a human source, the impeachment of machines is often more challenging. In the AI discipline of machine learning,[218] problems of intelligibility and explainability are becoming more perplexing as the discipline itself becomes more advanced. While the goal of increasing accuracy is an obvious pursuit, accuracy and explainability in machine learning have an inverse relationship.[219] As a result, algorithmic models touted for accuracy are often frustratingly opaque, giving rise to trade-offs between these two attributes and creating resultant challenges. For example, humans giving testimony to support the credibility of the machine may not know enough about what happens in the black box to speak authoritatively about it.

Because sophisticated algorithms are often not understandable by humans, they are unexplainable in court. Data scientists and machine learning practitioners providing testimony simply may not know enough about the training data[220] or the opaque

216. Kehl et al., *supra* note 76.

217. Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 1977 (2016).

218. James Furbush, *Machine Learning*, O'REILLY MEDIA (May 3, 2018), https://www.oreilly.com/content/machine-learning-a-quick-and-simple-definition/ [https://perma.cc/33A8-RFFF].

219. Jesus Rodriguez, *Interpretability vs. Accuracy: The Friction that Defines Deep Learning*, TOWARDS DATA SCI. (Jun 6, 2018), https://towardsdatascience.com/interpretability-vs-accuracy-the-friction-that-defines-deep-learning-dae16c84db5c [https://perma.cc/LF2C-BVYX].

220. Training data is also known as a training set or a training dataset. It is the initial dataset used to train an algorithm to make predictions.

functioning of the algorithm itself. As models get more sophisticated, this inverse relationship will widen. As Pasquale (2017) writes, outputs of black box algorithms "are secretly computed; they deny due process and intelligible explanations to defendants; and they promote a crabbed and inhumane vision of the role of punishment in society."[221]

Multiple proposals to implement statutory safeguards exist in literature, inclusive of calls from Roth (2008) for machine credibility testing and provisions for machine confrontation.[222] As Pasquale (2017) writes, "[a]t a bare minimum, governments should not use algorithms like the COMPAS score without some kind of external quality assurance enabled by qualified transparency."[223]

Proposed legislation is consistent with calls from legal scholarship. In September 2019, Rep. Mark Takano (D-CA) introduced the Justice in Forensic Algorithms Act of 2019[224] to "ensure that defendants have access to source code and other information necessary to exercise their confrontational and due process rights when algorithms are used to analyze evidence in their case."[225] This legislation includes a provision for third-party review, calling on the National Institute for Standards and Technology (NIST) to establish a Computational Forensic Algorithms Standards and a Computational Forensic Algorithms Testing Program, with a requirement to "address the potential for disparate impact across protected classes in standards and testing."[226]

2. Revising Rule Exceptions

In both criminal and civil cases, safeguards against algorithmic opaqueness exist in the form of evidence rules that treat machine testimony as an "expert witness", thus subjecting it to the *Daubert-Frye* standard(s)[227] that govern the admissibility of scientific evidence. One class of evidence routinely presented in criminal trials is DNA evidence, first used in a criminal case in the U.S. in 1987.[228] Today, all fifty states and the federal government require DNA samples to be collected from certain categories of offenders.[229] But even this class of probabilistic evidence is under new scrutiny, following the discovery that receipt of a bone marrow transplant can transform a

221. Frank Pasquale, *Secret Algorithms Threaten the Rule of Law*, MIT TECH. REV. (Jun 1, 2017), https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/ [https://perma.cc/W5BE-GB4J].

222. Roth, *supra* note 217, at 1978.

223. Pasquale, *supra* note 221.

224. Justice in Forensic Algorithms Act of 2019, H.R. 4368, 116th Congress (2019-2020) (available at https://www.congress.gov/bill/116th-congress/house-bill/4368/).

225. Press Release, Rep. Mark Takano (CA-41), Rep. Takano Introduces the Justice in Forensic Algorithms Act to Protect Defendants' Due Process Rights in the Criminal Justice System (Sep. 17, 2019).

226. *Id.*

227. *See generally* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 594 (1993); Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923) [hereinafter *Daubert-Frye*] (together these cases form the basis for the eponymous legal standard).

228. LAW DICTIONARY, *History of DNA Testing In Criminal Cases*, https://thelawdictionary.org/article/history-of-dna-testing-in-criminal-cases/ (last visited Jan. 21, 2021) [https://perma.cc/PU2A-9XXY].

229. U.S. DEP'T JUSTICE, *Advancing Justice Through DNA Technology: Using DNA to Solve Crimes*, https://www.justice.gov/archives/ag/advancing-justice-through-dna-technology-using-dna-solve-crimes (last visited Jan. 21, 2021) [https://perma.cc/9VHE-KDFY].

person into a chimera, a term that refers to someone with two sets of DNA.[230] While algorithmic predictions are generally not allowed as evidence in a trial, DNA notwithstanding, exceptions to evidence rules include other elements of criminal procedure in which algorithms are routinely used. For example, the Federal Rules of Evidence define exceptions for "miscellaneous proceedings, such as extradition or rendition; issuing an arrest warrant, criminal summons, or search warrant; a preliminary examination in a criminal case; sentencing; granting or revoking probation or supervised release; and considering whether to release on bail or otherwise."[231] Because of these exceptions, rules that exist to protect defendants from unfairness at the hands of algorithmic tools in a criminal trial offer no such protection pre-trial or post-trial. Furthermore, ninety-eight percent of criminal cases in the United States never go to trial and are plea-bargained,[232] rendering evidence rules minimally effective in ensuring due process and equal protection in the face of mathematical predictions.

Exceptions to evidence rules must be revisited and changed. Algorithmic instruments, although not used in the trial phase, are still a version of what Roth (2008) refers to as "machine evidence,"[233] and should thus be subject to the *Daubert-Frye*[234] standard(s). With the vast majority of criminal cases resolved without a trial, evidence rules in their current state do little to protect defendants from denial of due process or unfairness. As Roth (2008) argues, "[f]or machines offering 'expert' evidence on matters beyond the ken of the jury, lawmakers should clarify and modify existing *Daubert* and *Frye* reliability requirements for expert methods to ensure that machine processes are based on reliable methods and are implemented in a reliable way."[235]

A consistent theme is the need for greater openness and transparency. As Christopher Slobogin, director of the criminal justice program at Vanderbilt Law School, explains, "[r]isk assessments should be impermissible unless both parties get to see all the data that go into them. It should be an open, full-court adversarial proceeding."[236] Other proposals are more precisely targeted at specific elements of criminal procedure. McGarraugh (2013) calls for a proposed change to the Model Penal Code[237] to limit use of risk assessment instruments in sentencing, describing the change as a "reasonable compromise between fairness to defendants and protecting public safety" and arguing in favor of "establishing statutory criteria for determining if a risk assessment instrument is 'sufficiently reliable.'"[238] To the extent the U.S.

230. Heather Murphy, *When a DNA Test Says You're a Younger Man, Who Lives 5,000 Miles Away*, N.Y. TIMES (Dec. 7, 2019), https://www.nytimes.com/2019/12/07/us/dna-bone-marrow-transplant-crime-lab.html [https://perma.cc/A36L-VX6Q].

231. Fed. R. Evid. 1101(d)(3).

232. Gramlich, *supra* note 133.

233. Roth, *supra* note 217, at 1983.

234. *Daubert-Frye*, supra note 227.

235. Roth, *supra* note 217, at 1981-82.

236. Angwin et al., *supra* note 15.

237. Model Penal Code: Sentencing § 6B.09 (Tentative Draft No. 2, 2011).

238. McGarraugh, *supra* note 85, at 1081.

justice system is predicated on the presumption of innocence, the use of algorithmic tools to predict the probability of future crime in deciding the length of one's sentence is a contradiction.

3. Instituting Checks

Not all algorithms exist to simply make decisions on their own. They often serve the purpose of "decision support," providing humans with insights needed for more informed decisionmaking. For humans that rely on the output of machines to support decisions, the specter of automation bias[239] looms over the determination. Our human tendency to over-rely on a machine's judgment can have comical consequences, such as when a car's GPS system convinces its driver that the best route is to veer into the ocean.[240] But in a legal context, this tendency can have life-altering consequences for a defendant, whose future can hinge on the output of a computer algorithm and its unconscious influence over a human decisionmaker. As Citron writes, "[a]utomation bias effectively turns a computerized suggestion into a final, authoritative decision."[241]

Generally speaking, algorithms are trained by historical data in an attempt to bring objectivity to decisionmaking; but the biases embedded in data, and thus the system itself, can alter the behavior of its user. In describing use of a predictive model for child neglect and abuse, Eubanks (2017) observes user behavior in which "the algorithm seems to be training the workers."[242] It is imperative that users of these instruments, judges mostly, are subject to statutory or regulatory requirements to demonstrate understanding of their function and meaning. As Hannah-Moffat (2011) explains:

> My interviews with criminal justice practitioners demonstrated that few understand and appropriately interpret probability scores. Despite receiving training on these tools and their interpretation, practitioners tended to struggle with the meaning of the risk score and the importance of the items contained in the assessment tools.[243]

As noted earlier in this Section, literature on perceptions, attitudes, and usage of risk instruments among judges is lacking.

CONCLUSION

The criminal justice process consists of a chain of decisions that are each being increasingly influenced or automated with statistics and algorithms, creating a cascading effect that continues long after release and reentry. This is taking place despite

239. *See* Linda Skitka, Kathleen Mosier & Mark Burdick, *Does Automation Bias Decision-Making?*, 51 INT'L J. HUM.-COMPUT. STUD. 991 (1999)

240. Akiko Fujita, *GPS Tracking Disaster: Japanese Tourists Drive Straight into the Pacific*, ABC NEWS (Mar. 16, 2012), https://abcnews.go.com/blogs/headlines/2012/03/gps-tracking-disaster-japanese-tourists-drive-straight-into-the-pacific [https://perma.cc/DY5W-JM4P].

241. Danielle Citron, *(Un)Fairness Of Risk Scores In Criminal Sentencing*, FORBES (June 13, 2016), https://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-risk-scores-in-criminal-sentencing/#3a51a8944ad2 [https://perma.cc/6QRR-DKYY].

242. EUBANKS, supra note 97, at 142.

243. Hannah-Moffat, *supra* note 108, at 12 (omitting reference to studies).

a problematic lack of understanding about the complexity of the interaction between these decisions and the lack of transparency into the underlying algorithmic black boxes. Allowing one decision to influence the other without transparent understanding by all parties involved is a threat to fairness and due process. Collectively, calls from legal scholars, legislators, and social scientists brings much-needed attention to a troubled domain that is nonetheless eager to adopt new digital technology. The speed of adoption escalates the challenge, in that reforms such as elimination of cash bail[244] are taking place in parallel with enhancements to these instruments. The absence of a control variable makes it difficult to discern the net effect of either change.

This Article would be incomplete without a brief discussion of growing calls to ban the use of algorithmic tools altogether. In February 2020, the Pretrial Justice Institute (PJI) updated its position regarding the use of pretrial risk assessment tools, stating, "[w]e now see that pretrial risk assessment tools, designed to predict an individual's appearance in court without a new arrest, can no longer be a part of our solution for building equitable pretrial justice systems."[245] This was preceded by long-held positions by prominent legal scholars who have raised objections to varying degrees. Starr (2014) has offered consistent critiques of risk prediction instruments in sentencing, for example, arguing that "this practice violates the Equal Protection Clause and is bad policy: an explicit embrace of otherwise-condemned discrimination, sanitized by scientific language."[246] Responding to claims of incremental improvement to justify continued adoption, Huq (2019) writes:

> It is not sufficient . . . to point to a superseded technology that relies upon flawed human discretion and that already generates large racial effects as a justification for new, slightly less flawed technologies for allocating coercion. The mere fact that the status quo ante is characterized by racial injustice does not legitimatize proposals that preserve or extend some substantial part of that injustice.[247]

Given the pervasiveness of algorithmic tools with no uniform set of rules or guidelines for usage across the multitude of jurisdictions, a ban would be a more daunting challenge, but another valid proposal for reform nonetheless. Lawmakers are clearly paying close attention to the increasingly pervasive role of big data and algorithmic decisionmaking in our world. For example, the proposed Algorithmic Accountability Act[248] in the U.S. Congress strives to stem adverse and discriminatory commercial outcomes by asking the Federal Trade Commission to develop rules and guidelines to regulate automated decisionmaking for companies with access to large amounts of data on Americans. These consistent patterns and characteristics of how algorithms

244. Julie McMahon, *New York Ends Cash Bail For Most: What It Means For People Charged With A Crime*, SYRACUSE.COM (Apr. 2, 2019), https://www.syracuse.com/news/2019/04/new-york-ends-cash-bail-for-most-what-it-means-for-people-charged-with-a-crime.html [https://perma.cc/UFK5-C8FF].

245. PRETRIAL JUST. INST., *Updated Position on Pretrial Risk Assessment Tools* (Feb. 7, 2020), https://www.pretrial.org/wp-content/uploads/Risk-Statement-PJI-2020.pdf [https://perma.cc/NZ5Y-Z5U9].

246. Sonja Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 803 (2014).

247. Huq, *supra* note 8, at 1104.

248. Algorithmic Accountability Act of 2019, S. 1108, 116th Congress (2019-2020).

behave in commercial scenarios apply equally to how risk assessment algorithms are employed in the criminal justice process; but the challenges presented by the use of big data and algorithms in the legal domain persist, and the cascading effects of racial bias embedded in their inner workings can no longer be denied.