Connecticut Law Review   School of Law

2022

The Democratizing Potential of Algorithms?

Ngozi Okidegbe
Benjamin N. Cardozo School of Law

Article

The Democratizing Potential of Algorithms?

NGOZI OKIDEGBE

Jurisdictions are increasingly embracing the use of pretrial risk assessment algorithms as a solution to the problem of mass pretrial incarceration. Conversations about the use of pretrial algorithms in legal scholarship have tended to focus on their opacity, determinativeness, reliability, validity, or their (in)ability to reduce high rates of incarceration, as well as racial and socioeconomic disparities within the pretrial system. This Article breaks from this tendency, examining these algorithms from a democratization of criminal law perspective. Using this framework, it points out that currently employed algorithms are exclusionary of the viewpoints and values of the racially marginalized communities most impacted by their usage, since these algorithms are often procured, adopted, constructed, and overseen without input from these communities.

This state of affairs should caution enthusiasm for the transformative potential of pretrial algorithms since they reinforce and entrench the democratic exclusion that members from these communities already experience in the creation and implementation of the laws and policies shaping pretrial practices. This democratic exclusion, alongside social marginalization, contributes to the difficulties that these communities face in contesting and resisting the political, social, and economic costs that pretrial incarceration has had and continues to have on them. Ultimately, this Article stresses that resolving this democratic exclusion and its racially stratifying effects might be possible but requires shifting power over pretrial algorithms toward these communities. Unfortunately, the actualization of this prescription may be unreconcilable with the aims sought by algorithm reformers, revealing a deep tension between the algorithm project and racial justice efforts.

ARTICLE CONTENTS

The Democratizing Potential of Algorithms?

NGOZI OKIDEGBE [*]

INTRODUCTION

Bail decisions are among the most consequential decisions in our criminal legal system.[1] Their outcomes can lead to short-term or long-term incarceration before trial, enacting hardships on defendants, their families, and their communities.[2] Today, bail decisions are increasingly informed by pretrial algorithms that utilize an actuarial method and information about the defendant to determine the likelihood that the defendant will fail to appear

---

or be arrested for a pretrial crime if released before trial.[3] However, these pretrial algorithms have what I term an "input problem."[4] They inform life-altering decisions around pretrial release, detention, and electronic monitoring.[5] Yet, their employment and operation cannot be stopped, shaped, or influenced by the racially marginalized communities most likely to be impacted by these algorithms.[6]

The input problem has three dimensions. First, jurisdictions regularly adopt these algorithms in opaque ways without consulting racially marginalized communities, even though these communities are disproportionately affected by their utilization.[7] Second, these communities tend to be shut out of the algorithmic construction process,[8] meaning that the factors utilized by these algorithms have not been subjected to community scrutiny.[9] Moreover, these algorithms are often constructed by organizations

---

[3] Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2221–28 (2019) [hereinafter Mayson, *Bias*] ("Over the last five years, criminal justice risk assessment has spread rapidly."); Kia Rahnama, *Science and Ethics of Algorithms in the Courtroom*, 2019 U. ILL. J. L., TECH. & POL'Y 169, 171 (2019) ("These risk-assessment tools can match the information obtained from individual criminal defendants with the patterns observed among past offenders with a similar background and make probabilistic judgments about defendants' future conduct."). Though this Article focuses exclusively on the use of risk assessment algorithms in bail, there is a body of scholarship on their use in sentencing. *See, e.g.*, Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59 (2017) [hereinafter Eaglin, *Constructing*]; Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57 (2018); John Monahan, *Risk Assessment in Sentencing*, *in* 4 REFORMING CRIMINAL JUSTICE: PUNISHMENT, INCARCERATION, AND RELEASE 77, 78–79 (Erik Luna ed., 2017), https://law.asu.edu/sites/default/files/pdf/academy_for_justice/Reforming-Criminal-Justice_Vol_4.pdf

[4] The invocation of the term "input problem" contains a *double meaning* meant to capture both the fact that these algorithms utilize racially disparate and carceral inputs and that they are designed and implemented without input from the racially marginalized communities that stand to be most impacted by these tools' utilization.

[5] *See* Lauryn P. Gouldin, *Defining Flight Risk*, 85 U. CHI. L. REV. 677, 713 (2018) (noting that these risk assessment tools are employed to impact judicial determination in relation to "the calculation of nonappearance risk") ; Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 492–93 (2018) [hereinafter Mayson, *Dangerous*] ("To accomplish that objective, a growing number of jurisdictions are adopting actuarial risk-assessment tools to sort high-risk from low-risk defendants."); Chaz Arnett, *From Decarceration to E-carceration*, 41 CARDOZO L. REV. 641, 651 (2019) (noting that jurisdictions are employing risk assessment tools to "determine the degree of surveillance [that a defendant] should be subjected to if released").

[6] Eaglin, *Constructing*, *supra* note 3, at 88–99 (describing the lack of democratic input that goes into the construction and implementation of algorithmic tools used in the criminal legal system).

[7] Hannah Bloch-Wehba, *Visible Policing: Technology, Transparency, and Democratic Control*, 109 CALIF. L. REV. 917, 917–19 (2021) [hereinafter Bloch-Wehba, *Visible Policing*] (noting how the secrecy surrounding the adoption processes around algorithmic systems renders it difficult to track their deployment and use).

[8] Algorithmic construction concerns the process of data collection and selection, determination of the meaning of risk in the context of a measurable outcome in pretrial trial system, selection of predictive factors, and selection of risk thresholds. This process is generally performed by developers and other technocrats with no input from community members. For more information about algorithmic construction, see Eaglin, *Constructing*, *supra* note 3, at 73–88.

[9] Community groups have discussed this problem at length. *See, e.g.*, LEADERSHIP CONF. EDUC. FUND, THE USE OF PRETRIAL "RISK ASSESSMENT" INSTRUMENTS: A SHARED STATEMENT OF CIVIL RIGHTS CONCERNS 1–9, http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf (discussing lack of community consultation around algorithmic construction); Eaglin, *Constructing*, *supra* note 3, at 108 ("More often, however, the tools are developed by private entities and adopted by jurisdictions with limited opportunity for expert input and localized feedback.").

that have no internal mechanisms for facilitating outsider input about data collection and data utilization.[10] Third, even where the public is provided with an opportunity to express its views about impending pretrial algorithmic governance,[11] these participatory systems tend to be unresponsive to those who oppose this form of governance, particularly those hailing from communities that have been devastated by the carceral state. The combination of all of these practices results in an anti-democratic iteration of pretrial algorithmic governance that maintains these communities' traditional marginalization within pretrial governance.[12]

The input problem is neither exclusive to pretrial algorithms nor to racially marginalized communities. Indeed, a growing number of scholars, policymakers, and racial justice activists have raised serious concerns over the use of algorithms in governmental processes, especially since these algorithms are often deployed without meaningful public notice, input, or oversight.[13] This is partially caused by the fact that jurisdictions have tended to outsource the construction, implementation, and ongoing maintenance of such algorithms to the private sector.[14] This privatization has enabled algorithms to escape meaningful public debate and oversight, to the detriment of traditional accountability metrics.[15]

But conversations about the absence of public input around algorithmic governance have largely failed to attend to the particular harms that the input problem imposes on poor and racially marginalized people.[16] For these communities, the input problem goes beyond the fact that algorithmic governance, as currently employed, is opaque or non-inclusive of the oppressed populations that are most likely to interact with it—though both

---

[10] Eaglin, *Constructing*, *supra* note 3, at 118–21.

[11] This Article's definition of "pretrial algorithmic governance" is informed by and adds to Hannah Bloch-Wehba's definition of algorithmic governance, which is "the use of automated decision-making methodologies by governments to inform the policymaking and adjudicative process." Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1267 (2020) [hereinafter Bloch-Wehba, *Algorithms*].

[12] This Article uses the term "pretrial governance" to refer to the creation and implementation of pretrial laws, policies, and practices.

[13] *See, e.g.*, Kate Weisburd, *Punitive Surveillance*, 108 VA. L. REV. (forthcoming 2022) (raising concerns about the use of electronic surveillance); Bloch-Wehba, *supra* note 11 (raising concerns around the opacity in the adoption and deployment of algorithmic systems; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018) (noting the transparency concern around the use of AI-systems); Barry Friedman & Maria Ponomarenko, *Democratic Policing*, 90 N.Y.U. L. REV. 1827 (2015) (problematizing the lack of democratic input around police use of surveillance technologies).

[14] Bloch-Wehba, *Algorithms*, *supra* note 11, at 1272; Kate Crawford & Jason Schultz, *AI Systems as State Actors,* 119 COLUM. L. REV. 1941, 1942 (2019).

[15] *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 10–15 (2015) (noting how algorithmic processes are opaque to everyday citizens, decisionmakers, and legal processes).

[16] *But see* Dorothy E. Roberts, *Digitizing the Carceral State*, 132 HARV. L. REV. 1695 (2019) [hereinafter Roberts, *Digitizing*] (reviewing VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018)); Jessica M. Eaglin, *Technologically Distorted Conceptions of Punishment*, 97 WASH. U. L. REV. 483 (2019) [hereinafter Eaglin, *Technologically*] (recognizing the racial dimension of the use of risk assessment tools)

issues pose barriers to rendering algorithms accountable to these communities.[17] The problem is that the exclusion of these communities within algorithmic governance reinforces and legitimizes the barriers that already impede their ability to challenge or to gain control over the criminal legal institutions responsible for their oversurveillance, overcriminalization, and over-incarceration.[18]

In the pretrial algorithmic governance context, these communities' exclusion has already produced tangible and material consequences. Broadly speaking, three consequences flow from their exclusion. First, their exclusion from algorithmic construction has enabled the production of pretrial algorithms that maintain existing racial disparities in the pretrial system due to their utilization of inputs that are racially disparate, carceral, and fail to account for the individual and communal harms that pretrial incarceration enacts.[19]

Second, their exclusion within pretrial algorithmic governance operates as an additional barrier to their efforts to resist a pretrial system that overincarcerates their community members to the detriment of their communal safety and cohesion.[20] Third, the combination of the first two consequences threatens to "lock-in"[21] the racial status quo and resultingly hampers the ability of these communities to collectively contest the political,

---

[17] It is important to note that algorithmic governance by private actors did not inaugurate this exclusion. Rather, this exclusion is a hallmark of pretrial governance, which itself operates to mute the viewpoints and values of these communities. *See* Jocelyn Simonson*, Democratizing Criminal Justice Through Contestation and Resistance*, 111 Nw. U. L. REV. 1609 (2017) [hereinafter Simonson, *Democratizing Criminal Justice*].

[18] For more information regarding the particular harms that the criminal legal system enacts on Black people, see Devon W. Carbado, *Blue-on-Black Violence: A Provisional Model of Some of the Causes*, 104 GEO. L.J. 1479, 1485 (2016) (describing the vulnerability that Black people experience at the hands of criminal legal institutions through repeated police interactions, which include policing practices, mass criminalization, racial stereotyping, and racial segregation); Paul Butler, *The System Is Working the Way It Is Supposed To: The Limits of Criminal Justice Reform*, 104 GEO. L.J. 1419, 1423–25 (2016) (describing how the racial injustices experienced by Black people are the objective—as opposed to the unintended consequence of the criminal legal system's operation); Jamelia N. Morgan, *Rethinking Disorderly Conduct*, 109 CALIF. L. REV. 1637 (2021) (discussing the intersection of race and disability in the criminalization, surveillance, and incarceration of bodies as it pertains to the enforcement of disorderly conduct offenses); Michael Pinard, *Race Decriminalization and Criminal Legal System Reform*, 95 N.Y.U. L. REV. ONLINE 119, 120 (2020) ("While the criminal legal system particularly infiltrates the lives of poor people, it is singularly relentless and merciless on Black men, women, and children. It is common knowledge that Black communities have borne the brunt of mass incarceration, mass convictions, and every other aspect of the criminal legal system.").

[19] Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 806 (2014) (contending that neutral inputs can operate as proxies for race that "can be expected to contribute to the concentration of the criminal justice system's punitive impact among those who already disproportionately bear its brunt, including people of color").

[20] The adverse effects of this exclusion have pushed some community groups to push for community representation in algorithmic governance. *See* CMTY. JUST. EXCH., AN ORGANIZER'S GUIDE TO CONFRONTING PRETRIAL RISK ASSESSMENT TOOLS IN DECARCERATION CAMPAIGNS 33 (2019), https://perma.cc/YLC4-MR4H (discussing the ways in which community members can influence and shape the creation and deployment of algorithms used in the criminal legal system).

[21] Rebecca Crootof, *"Cyborg Justice" and the Risk of Technological-Legal Lock-In*, 119 COLUM. L. REV. F. 233, 235 (2019) ("Translating rules and decisionmaking procedures into algorithms grants them a new kind of permanency, which creates an additional barrier to legal evolution.").

economic, and social costs that mass pretrial incarceration, alongside other carceral practices, has had and continues to have on their communities.[22] The extent of these consequences should dampen enthusiasm about the transformative potential of algorithms since, unless the input problem is attended to, well-meaning algorithmic reforms designed to redress racial inequity in the pretrial system are doomed to reproduce existing barriers to racial justice.[23] This result should raise questions about the compatibility of algorithmic governance with efforts to challenge and resist the racial and socioeconomic bias and disparity of the current pretrial system and beyond.

This Article explores the consequences of the input problem for racially marginalized communities and the difficulties associated with its redress. Using the example of democratic exclusion in pretrial algorithmic governance, its central claim is that the input problem limits the capacity of pretrial algorithmic governance to combat racial injustices within the pretrial system, since it contributes to the democratic exclusion experienced by these oppressed communities in pretrial governance. This democratic exclusion, in turn, works to deny members of these communities the agency to collectively resist mass pretrial incarceration and pursue their own visions of public safety within the pretrial context. Redressing the input problem might be possible, but it requires a meaningful shift in power over pretrial algorithmic governance to oppressed groups and, in so doing, endowing them with control over if and on what basis algorithmic-based reforms are pursued. But the actualization of this prescription would be in tension with the intended aims and objectives sought by proponents of algorithm-based reforms ("algorithm reformers"). On the other hand, more modest approaches to redressing the input problem risk largely preserving it while enabling pretrial algorithmic governance to benefit from the veneer of communal approval. Given this, this Article concludes that both the extent of change required to attend to the input problem and the potential backlash in so attending reveal a deep, unreconcilable tension between the algorithm project and racial justice efforts.

This Article makes three contributions to the existing literature. First, it adds to the current scholarly conversation on algorithms, accountability, and community participation. Many scholars have explored how the opaque ways in which algorithms are procured and designed obstruct efforts to subject automated governmental decision-making to public scrutiny and

---

[22] *See* Dorothy E. Roberts, *Democratizing Criminal Law as an Abolitionist Project*, 111 Nw. U. L. Rev. 1597, 1604 (2017) [hereinafter Roberts, *Democratizing*] (noting that the system of incarceration "targets entire black communities for state regulation that deprives them of the resources, liberties, and legitimacy needed for democratic participation").

[23] *See* Jocelyn Simonson, *Police Reform Through a Power Lens*, 130 Yale L.J. 778, 807 (2021) [hereinafter Simonson, *Police Reform*] (discussing this problem in policing process and contending that layers of democratic exclusion in the criminal legal system have facilitated the "reproducing and legitimizing [of] an unequal and racialized system of justice").

oversight.[24] Yet, these discussions tend to revolve around transparency, without sufficient attention to power.[25] Though transparency is a crucial precondition to rendering algorithmic governance democratically accountable to the public,[26] transparency alone cannot attend to the multiple layers of democratic exclusion experienced by oppressed communities. Transparency may redress power inequities stemming from informational discrepancies existing between government institutions and the public generally.[27] It may also increase public participation in algorithmic governance. However, transparency on its own is inattentive to the "layers of democratic exclusion" that reinforce the political powerlessness experienced by those most harmed by the system, who are thus unable to change the system or dismantle and reconstitute it.[28] By attending to power, this Article expands the contours of the traditional democratic critique in algorithmic literature and is the first to connect it to the burgeoning body of criminal procedure scholarship considering the promises and pitfalls of democratizing criminal legal institutions for oppressed communities.[29]

Second, this Article joins the racial justice strands of the democratization of criminal law scholarship[30] with race critical code

---

[24] *See* Bloch-Wehba, *Visible Policing*, *supra* note 7, at 919 (discussing how "new surveillance technology tends to operate in opaque and unaccountable ways, augmenting police power while remaining free of meaningful oversight); Andrew Guthrie Ferguson, *Illuminating Black Data Policing*, 15 OHIO ST. J. CRIM. L. 503, 504 (2018) (noting that "big data policing is opaque, lacking transparency because most of the magic happens as a result of "black box" proprietary and mathematically complex algorithms"). Amanda Levendowski, *Resisting Face Surveillance with Copyright Law*, N.C.L. REV. (forthcoming 2022) (noting the secrecy around the datasets used by facial surveillance companies)

[25] Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 615 (2019) (contending that "[u]sing algorithms to make commercial and social decisions is really a story about power, the people who have it, and how it affects the rest of us"). *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1258 (2008) (contending that automation transparency is insufficient); Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1024 (2017) (contending that algorithmic transparency is insufficient to redress algorithmic discrimination because "[e]ven a transparent, facially neutral algorithm can still produce discriminatory results").

[26] It is also important to note that, for those seeking to subvert existing technologies for racial justice aims, transparency is critical. *See, e.g.*, Bennett Capers, *Policing, Technology, and Doctrinal Assists*, 69 FLA. L. REV. 723, 725 (2017) (contending that techno-policing, such as body cams, could be harnessed to remake and enhance Fourth Amendment protections).

[27] Bloch-Wehba, *Visible Policing*, *supra* note 7, at 922–23.

[28] Simonson, *Democratizing Criminal Justice*, *supra* note 17, at 1610.

[29] Simonson, *Police Reform*, *supra* note 23, at 799–802; K. Sabeel Rahman & Jocelyn Simonson, *The Institutional Design of Community Control*, 108 CALIF. L. REV. 679, 696–99 (2020) (advocating for endowing oppressed communities with control over governmental institutions involved in racial and class subordination).

[30] This Article engages with a specific strand of the democratization of criminal law scholarship, which is concerned with redressing the democratic harms of the criminal legal system that subordinates oppressed communities in our society. In this context, democratization involves endowing these communities with a level of control over the criminal legal system in order to eliminate these harms. *See* Roberts, *Democratizing*, *supra* note 22, at 1605 ("Democratizing criminal law requires dismantling itsanti-democratic aspects altogether and reconstituting the criminal justice system without them.")

scholarship[31] to reveal and to critique the technological determinist[32] undercurrent in current legal scholarship on algorithmic governance and democracy. Thus far, the race critical code literature and the democratization of criminal law literature have largely evolved on separate tracks. Tapping into both sets of literature, this Article rejects the dominant technological determinist framing in which algorithmic governance is presented as inaugurating the problem of democratic exclusion. By so doing, this Article argues that algorithmic governance serves merely to entrench and to legitimate the existing democratic exclusion experienced by racially marginalized people in the crafting and implementation of criminal laws and policies. This recognition permits us to start to grapple with the ways in which new technologies reinforce the racial status quo, augmenting, as opposed to creating, the political powerlessness of already disempowered communities to contest and to resist their subordination within the criminal legal system.[33]

Third, this Article joins a growing set of scholarship considering the congruency between algorithm-based reforms and racial justice efforts. These scholars question the viability of repurposing algorithmic tools in order to challenge and contest structures of power.[34] This Article adds

---

[31] *E.g.*, RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE 9 (2019) [hereinafter BENJAMIN, ABOLITIONIST] (noting how "race itself [is] a kind of tool—one designed to stratify and sanctify social injustice as part of the architecture of everyday life"); SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS 11 (2015) [hereinafter BROWNE, DARK MATTERS]; SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM 1 (2018); Ruha Benjamin, *Introduction: Discriminatory Design, Liberating Imagination*, *in* CAPTIVATING TECHNOLOGY: RACE, CARCERAL TECHNOSCIENCE, AND LIBERATORY IMAGINATION IN EVERYDAY LIFE 1–22 (Ruha Benjamin ed., 2019) [hereinafter Benjamin, *Discriminatory Design*].

[32] I use the term "technological determinism" as invoked in TIM JORDAN, HACKING: DIGITAL MEDIA AND TECHNOLOGICAL DETERMINISM 13 (2008) to refer to the belief that social structures and values are highly shaped and augmented by new technologies. *See also* BENJAMIN, ABOLITIONIST, *supra* note 31, at 40; BROWNE, DARK MATTERS, *supra* note 31, at 8–9 (cautioning, in the context of surveillance, against viewing "surveillance as something inaugurated by new technologies," and instead seeing it as ongoing, which "is to insist that we factor in how racism and antiblackness undergird and sustain the intersecting surveillances of our present order").

[33] Benjamin, *Discriminatory Design*, *supra* note 31, at 3 (discussing how "emerging technologies can reinforce interlocking forms of discrimination, especially when we presume they are insulated from human influence").

[34] *See, e.g.*, Sean Allan Hill II, *Bail Reform & the (False) Racial Promise of Algorithmic Risk Assessment*, UCLA L. REV. (forthcoming 2022) (contending that pretrial algorithms are incompatible with achieving racial justice aims in the pretrial system); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1055–56 (2019) (contending that the racial equity of algorithms depends on their impact on existing racial stratification); SASHA COSTANZA-CHOCK, DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED 63 (2020) (noting that "[a] prison abolitionist stance does not support allocating additional resources to the development of tools [such as risk assessment] that extend the [Prison Industrial Complex], even to make them 'less biased'"); BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 2–3 (2007) (discussing the problems with risk prediction in criminal legal context); BENJAMIN, ABOLITIONIST, *supra* note 31, at 52–54 (raising doubts about the viability of achieving anti-racist ends by utilizing currently employed technological systems). *But cf.* Andrew Guthrie Ferguson, *The Exclusionary Rule in the Age of Blue Data*, 72 VAND. L. REV. 561, 630–31 (2019) (contending predictive technologies should be used to monitor and to check police behavior); Vincent M. Southerland, *The Intersection of Race and Algorithmic Tools in the Criminal Legal System*, 80 MD. L. REV. 487, 494, 532 (2021) (proposing that predictive technologies should be used for abolitionist projects).

another layer to this critique by advancing that the solvability of the democratic exclusion problem caused by algorithmic governance will affect the usefulness of algorithm-based reforms by those engaged in anti-racist projects. If algorithms cannot endow the most impacted communities with a voice to direct a system that has traditionally subordinated and muted their interests, then it is of no use to those in the struggle for racial justice.

This Article proceeds in five parts. Part I contextualizes the rise of risk assessment algorithms in the pretrial process and explains their methodology. Part II details the input problem and explains how currently employed algorithms reproduce and maintain the democratic exclusion experienced by marginalized communities within pretrial governance and beyond. Part III theorizes potential approaches to redressing the input problem. It then sets out a power-shifting model for combating the input problem: the creation of bail commissions at the state and federal levels consisting of members from affected communities who have the power to determine if and on what basis to pursue pretrial algorithmic governance. Part IV addresses the benefits of this approach. Part V sets out anticipated objections to actualizing this approach and how these objections signal a potential incompatibility of the algorithm project with democratizing efforts that are rooted in racial justice.

Two caveats are in order. First, this Article is primarily focused on the tension associated with redressing the input problem. It makes no claim about the outcomes that might be achieved through shifting power over algorithmic governance toward historically disempowered groups. As many scholars have observed in their explorations of endowing politically and historically marginalized people with control over criminal legal institutions, shifting power guarantees no particular outcome.[35] It could spell the end of algorithmic governance or reproduce the same or a more punitive and racially disparate version of it. The point of surfacing the input problem is to facilitate a greater appreciation for the racial justice implications and first-principle disagreements about the continuation of algorithmic governance. Second, in critiquing the lack of power and agency that low-income, racially marginalized communities have in algorithmic governance, this Article makes no claim about the role that developers, criminologists, policymakers, and other technocrats should play in algorithmic governance if the power-shifting model is enacted. Shifting power does not mean the end of reliance on the expertise possessed by technocrats; rather, it spells the dismantlement of the exclusive reliance on this form of expertise. The expected backlash to such a dismantlement reveals one of the several tensions associated with redressing the input problem, demonstrating this Article's central claim.

---

[35] Simonson, *Police Reform*, *supra* note 23, at 807.

I. Contextualization

This Part contextualizes the proliferation of pretrial algorithmic governance. It then sets out the methodology behind currently employed algorithms, setting the stage for the discussion of the input problem.

A. *The Rise of Pretrial Algorithmic Governance*

Pretrial algorithmic governance has become a popular component of recent efforts to reform the pretrial system.[36] Approximately twenty-five percent of Americans live in a jurisdiction that uses a pretrial algorithm.[37] Nearly every state has at least one county that uses a pretrial algorithm.[38] In the last decade, at least twenty-five states have either implemented or tabled legislation mandating the use of pretrial algorithms statewide[39]—even New York state, which passed sweeping legislation to curtail the practice of conditioning pretrial release on cash bail for most nonviolent offenses, made allowances for the continued use of algorithms in the pretrial setting to determine release conditions.[40] The trend towards pretrial algorithmic governance appears set to continue, as political pressure to depopulate jails mounts in the wake of the COVID-19 pandemic.[41]

The expanded reliance upon pretrial algorithmic governance has arisen from a growing recognition that our current bail system is plea-inducing,[42]

---

[36] Shaila Dewan & Carl Hulse, *Republicans and Democrats Cannot Agree on Absolutely Anything. Except This.*, N.Y. TIMES (Nov. 14, 2018), https://www.nytimes.com/2018/11/14/us/prison-reform-bill-republicans-democrats.html.

[37] PRETRIAL JUST. INST., THE STATE OF PRETRIAL JUSTICE IN AMERICA 13 (2017) ("The good news is that this analysis shows 25% of people living in the United States now reside in a jurisdiction that uses a validated evidence-based pretrial assessment.").

[38] *Where Are Risk Assessments Being Used?*, MAPPING PRETRIAL INJUSTICE, https://pretrialrisk.com/national-landscape/where-are-prai-being-used/ (last visited Jan. 8, 2021).

[39] Alicia Solow-Niederman, YooJung Choi & Guy Van den Broeck, *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L.J. 705, 714 (2019) ("In the last seven years alone, half of U.S. states have either implemented or are seriously considering the use of some form of risk assessment tools in pretrial settings.").

[40] FY 2020 NEW YORK STATE EXECUTIVE BUDGET: PUBLIC PROTECTION AND GENERAL GOVERNMENT: ARTICLE VII LEGISLATION 182 (2019) [hereinafter New York Bail Reform Law], https://www.budget.ny.gov/pubs/archive/fy20/exec/artvii/ppgg-artvii.pdf .

[41] Rafael A. Mangual, *Not Quite Fixed: Adjustments to New York's Bail Reform Leave Many of the Law's Holes Unplugged*, CITY J. (Apr. 16, 2020), https://www.city-journal.org/risks-of-bad-bail-reform (advocating for the greater use of risk assessment algorithms in the pretrial system in the wake of COVID).

[42] Megan Stevenson & Sandra G. Mayson, *Pretrial Detention and Bail*, *in* 3 REFORMING CRIMINAL JUSTICE: PRETRIAL AND TRIAL PROCESSES 21, 22 (Erik Luna ed., 2017) [hereinafter Stevenson & Mayson, *Pretrial Detention*], https://law.asu.edu/sites/default/files/pdf/academy_for_justice/Reforming-Criminal-Justice_Vol_3.pdf ("The increase in convictions [following pretrial detention] is primarily an increase in guilty pleas among defendants who otherwise would have had their charges dropped. The plea-inducing effect of detention undermines the legitimacy of the criminal justice system itself—especially if some of those convicted are innocent.").

costly,[43] and racially[44] and socioeconomically[45] disparate. Nearly 500,000 defendants in jail are pretrial detainees.[46] This means that sixty-six percent of inmates in city and county jails have not been convicted of a crime but instead are awaiting trial and the judicial resolution of their case.[47] Moreover, the pretrial detainee population is racially distorted. To illustrate this, it is useful to consider the racial composition of inmates held in city and county jails, since the majority of this population is pretrial detainees. Black people constitute thirty-three percent of jail inmates, despite constituting only thirteen percent of the population.[48] The jail incarceration rate for Black people (592 per 100,000)[49] is more than three times that of white people (187 per 100,000).[50]

These figures have fueled momentum around pretrial reform. In theory, the pretrial system seeks to balance a defendant's liberty interest before trial with the societal interest in the incapacitation of dangerous defendants and the adjudication of criminal offenses.[51] Its purpose is to release defendants before trial, except where their release poses a risk of non-appearance, pretrial crime, or obstruction of justice.[52] In practice, the pretrial process differs from this ideal. First, a defendant's pretrial detention is more predicated on ability to afford the assigned bail bond than the risk of flight or crime.[53] Second, judges often require bail, even though in most cases it is unnecessary to ensure a defendant's court appearance and law-abiding behavior.[54] For instance, in bail schedule jurisdictions, bail judges are mandated to set a predetermined monetary bail amount solely based on the

---

[43] *Id.* ("Pretrial detention has profound costs. In fiscal terms, the total annual cost of pretrial jail beds is estimated to be $14 billion, or 17% of total spending on corrections.").

[44] Yang, *supra* note 2, at 1465–67.

[45] COLIN DOYLE, CHIRAAG BAINS & BROOK HOPKINS, CRIM. JUST. POL'Y PROGRAM, BAIL REFORM: A GUIDE FOR STATE AND LOCAL POLICYMAKERS 12 (2019), https://university.pretrial.org/viewdocument/bail-reform-a-guide-for-state-and ("This means that a defendant's release depends upon an ability to pay. Wealthy defendants walk free while poor defendants languish in jail.").

[46] ZHEN ZENG, BUREAU OF JUST. STAT., U.S. DEP'T OF JUST., NCJ 253044, JAIL INMATES IN 2018 1 (2020), https://www.bjs.gov/content/pub/pdf/ji18.pdf.

[47] *Id.* at 6.

[48] *Id.* at 6, 12.

[49] *Id.* at 4, 13.

[50] *Id.*

[51] Stevenson & Mayson, *Pretrial Detention*, *supra* note 42, at 24; Yang, *supra* note 2, at 1416.

[52] Stevenson & Mayson, *Pretrial Detention*, *supra* note 42, at 24.

[53] Megan T. Stevenson, *Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes*, 34 J.L., ECON., & ORG. 511, 513 (2018) (arguing that "most people who are detained pretrial are detained due to an inability to pay bail").

[54] Bail is set in most cases because of the assumption that bail is required to ensure a defendant's appearance or compliance with the law on release. This assumption does not bear out in practice. *See* DOYLE, BAINS & HOPKINS, *supra* note 45, at 12–13 ("Money bail's connection to public safety is tenuous at best. Bail is not a means of preventing or deterring a defendant from committing crimes before trial. . . . Money bail is not necessary to ensure that defendants reappear for trial."). It is also important to note the importance of community bail funds in undermining the necessity of money bail to prevent pretrial crime and flight. *See* Jocelyn Simonson, *Bail Nullification*, 115 MICH. L. REV. 585, 591 (2017) [hereinafter Simonson, *Nullification*] ("Community bail funds provide to the public real-life examples of indigent defendants returning to court without having undermined public safety, despite an expert judicial determination that personal money was needed to prevent flight and mayhem.").

crime charged,[55] an amount that bears no relation to a defendant's flight or crime risk.[56] Similarly, in jurisdictions where bail is set on the basis of statutory factors, bail determinations are often rushed and conducted without sufficient attention to the full range of nonfinancial release options available to manage a defendant's pretrial risk.[57]

The greater utilization of pretrial algorithmic governance is one of dozens of bail reform proposals percolating around the country.[58] Pretrial algorithms are central to this project, and they tend to be risk assessment algorithms that use an actuarial method and information about the defendant to predict the likelihood that the defendant will fail to appear or will be arrested for pretrial misconduct if released before trial.[59] The risk prediction or "risk score" produced is then used as a factor for detaining the defendant or for determining "the degree of surveillance [the defendant] should be subjected to if released."[60] This measure aims to improve bail decision-making by conditioning release and detention on a defendant's riskiness to public safety.[61] The technology is intended to empower bail judges to identify and release low-risk defendants and to reserve pretrial detention for high-risk defendants.[62] The hope behind this technology is that bail judges will rely on the assessments supplied by these algorithms rather than their own subjective views of a defendant's riskiness—views that are often riddled with inaccuracy, irrationality, and bias.[63] Empirical studies demonstrate that bail decisions are susceptible to racial bias. For instance, a 2017 study by David Arnold, Will Dobbie, and Crystal Yang, which compared the pretrial conduct of marginally released white defendants with marginally released Black defendants, found that bail judges erroneously overestimated Black defendants' pretrial crime risk in comparison to that of

---

[55] Samuel R. Wiseman, *Fixing Bail*, 84 GEO. WASH. L. REV. 417, 445 (2016).

[56] In recent years, bail schedules have been the subject of constitutional challenges under the Due Process and Equal Protection Clauses for discriminatorily tying pretrial release to monetary amounts unaffordable to indigent defendants without a legitimate or compelling reason. *See, e.g.*, ODonnell v. Harris Cnty., 251 F. Supp. 3d 1052, 1067 (S.D. Tex. 2017), *appeal filed*, No. 17-20333 (5th Cir. Feb. 14, 2018).

[57] Cynthia E. Jones, *"Give Us Free": Addressing Racial Disparities in Bail Determinations*, 16 N.Y.U. J. LEGIS. & PUB. POL'Y 919, 932 (2013).

[58] Baughman, *supra* note 1, at 949, 981–85 (noting the various types of bail reform currently underway in the country); Ben Grunwald, *How to Reduce the Prison Population by X%* (unpublished manuscript) (on file with author) (contending that eliminating bail is one option for reducing prison populations).

[59] Mayson, *Dangerous*, *supra* note 5, at 509 ("Statisticians develop such tools by analyzing aggregated pretrial data to identify the traits of defendants that correlate most closely with the outcome of concern. Those traits are deemed 'risk factors.' The developers then create a checklist that assigns each risk factor a number of points corresponding to how closely it is correlated with the bad outcome in the group data.") (footnotes omitted).

[60] Arnett, *supra* note 5, at 651.

[61] Mayson, *Dangerous*, *supra* note 5, at 492–93 ("The core reform goal is to untether pretrial detention from wealth and tie it directly to risk. To accomplish that objective, a growing number of jurisdictions are adopting actuarial risk-assessment tools to sort high-risk from low-risk defendants.").

[62] *Id.* at 515.

[63] *Id.* at 508, 516.

white defendants, whose risk was correspondingly underestimated.[64] More specifically, in their evaluation of 177 bail judges, the researchers found that marginally released white defendants were twenty-two point two percent more likely to be rearrested for alleged pretrial misconduct than were marginally released Black defendants.[65] This variance could not be explained by white-Black differences in the characteristics (such as criminal history) or the type of crimes charged among the study participants.[66] Instead, the study concluded that the higher rate in arrests for offenses allegedly committed by marginally released white defendants demonstrated that bail judges were "relying on inaccurate stereotypes that exaggerate the relative danger of releasing [B]lack defendants versus white defendants at the margin."[67] In so finding, their study is reflective of prior studies that have been conducted throughout the country, evidencing the extent to which racial stereotypes infect and taint the bail determination process.[68]

The increased use of pretrial algorithms has ignited a fraught debate in bail reform circles regarding the merits of using predictive technologies in the pretrial process. Algorithm reformers claim that algorithms offer an objective and evidence-based path towards substantially lowering high rates of incarceration[69] and racial and socioeconomic disparities without endangering community safety, by identifying for pretrial release those posing a low risk of non-appearance and arrest for offending.[70]

---

[64] David Arnold, Will Dobbie & Crystal S. Yang, *Racial Bias in Bail Decisions*, 133 Q.J. ECON. 1885, 1917 (2018).

[65] *Id.* (noting that "[t]aken together, these [results] imply that marginally released white defendants are 22.2 percentage points more likely to be rearrested prior to disposition than marginally released black defendants . . . consistent with racial bias against [B]lacks").

[66] *Id.* at 1929 (stating that "[o]ur estimates are nearly identical if we account for observable crime and defendant differences by race, indicating that our results cannot be explained by [B]lack–white differences in the probability of being arrested for certain types of crimes (e.g., the proportion of felonies versus misdemeanors) or [B]lack–white differences in defendant characteristics (e.g., the proportion of defendants with prior offenses versus no prior offenses)").

[67] *Id.*

[68] *See, e.g.*, Meghan Sacks, Vincenzo A. Sainato & Alissa R. Ackerman, *Sentenced to Pretrial Detention: A Study of Bail Decisions and Outcomes*, 40 AM. J. CRIM. JUST. 661, 677–78 (2015) (discussing the issue of racial difference in imposition of monetary bail amounts).

[69] Christopher Slobogin, *Preventive Justice: How Algorithms, Parole Boards and Limiting Retributivism Could End Mass Incarceration*, 56 WAKE FOREST L. REV. 97 (2021) ("[T]he quantified results of well-validated [risk assessment instruments] can provide a concrete, rational basis for diversion or release. If . . . adherence to those results is *required* in most circumstances, the human urge to incapacitate those in the law's grasp can be even more effectively resisted because decision-makers must obey the objective facts.").

[70] *See, e.g.*, Stevenson & Mayson, *Pretrial Detention*, *supra* note 42, at 29–34 (describing the how the turn to actuarial risk assessment holds the promise of reducing unwarranted disparities); Sam Corbett-Davies, Sharad Goel & Sandra González-Bailón, *Even Imperfect Algorithms Can Improve the Criminal Justice System*, N.Y. TIMES (Dec. 20, 2017), https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html (describing the benefits that can accrue from using algorithms in the criminal legal system); Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 875 (2016) (describing one of the allures of machine learning algorithms as its ability to ostensibly disaggregate low-risk defendants from high risk ones even where the factors that tend to increase a defendant's probability of committing misconduct is multifaceted); Mirko Bagaric & Gabrielle Wolf, *Sentencing by Computer: Enhancing Sentencing*

At the same time, a racial justice movement has coalesced in opposition to the use of pretrial algorithmic governance, voicing concerns about its bias,[71] fairness,[72] due process,[73] and opacity.[74] They fear that the increased employment of these algorithms will maintain the racial status quo within the pretrial system, while simultaneously obscuring—or worse, legitimizing—its existence.[75] The above racial justice concerns have begun to materialize.[76] Part II will continue to discuss ways in which the input problem has contributed to the racial justice challenge presented by algorithmic governance.

The next section of this Article introduces the kind of algorithms used in the pretrial process. Its aim is not to provide a comprehensive overview of the inner workings of all the pretrial algorithms that have been or are being developed in the country.[77] Instead, this Part examines the common features of these algorithms, setting the stage for the discussion in Part II about how the

---

Transparency and Predictability, and (Possibly) Bridging the Gap Between Sentencing Knowledge and Practice, 25 GEO. MASON L. REV. 653, 654 (2018) (contending another appeal of algorithmic decision-making is its promise of rendering consistent and predictable decisions).

[71] Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1671–72 (2020).

[72] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 672 (2016); Ion Meyn, *Race-Based Remedies in Criminal Law*, 63 WM. & MARY L. REV. 219, 247–51 (2021).

[73] Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1375–76 (2018); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 663 (2018).

[74] Bloch-Wehba, *Algorithms*, *supra* note 11, at 1267 (noting that the opacity of algorithmic governance is a barrier to subjecting algorithms to public scrutiny and debate). *But see* Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 114 (2018) (contending that "[t]he opacity of the algorithm does not prevent us from scrutinizing its construction").

[75] *See, e.g.*, Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947, 982 (cautioning that moving toward big data algorithmic solutions risks further entrenching past racial inequities while making them "harder to successfully challenge and expose because they are presented as part of the 'hard science' of big data"); Eaglin, *Technologically*, *supra* note 16, at 487 (noting how sentencing risk assessment tools function to "reify structural racism under the auspice of scientific objectivity.").

[76] The greater use of this technology has not substantially reduced racial disparities in the pretrial process though it has reduced overall number of persons detained pretrial trial. The experience in New Jersey is illustrative. In 2017, New Jersey implemented PSA statewide, overhauling its cash-based system in favor of a system of detention based on risk. GLENN A. GRANT, N.J. JUDICIARY, CRIMINAL JUSTICE REFORM: JAN 1–DEC 31 2018 REPORT TO THE GOVERNOR AND THE LEGISLATURE 3–4 (2019), https://njcourts.gov/courts/assets/criminal/2018cjrannual.pdf. The switch was lauded as a success, with the pretrial population falling by nineteen percent in its first year and by thirteen percent in its second year. *Id.* at 38. The racial makeup of the New Jersey pretrial population has largely remained constant, despite the reform. *Id.* at 27 (noting that the racial percentage of Blacks in jail in 2018 was the same as 2012; however, between this period, the Latinx population declined by two percent and the white population increased by two percent). A similar result occurred in North Carolina. CINDY REDCROSS, BRIT HENDERSON, LUKE MIRATRIX & ERIN VALENTINE, MDRC CTR. FOR CRIM. JUST. RSCH., PRETRIAL JUSTICE REFORM STUDY: EVALUATION OF PRETRIAL JUSTICE SYSTEM REFORMS THAT USE THE PUBLIC SAFETY ASSESSMENT: EFFECTS IN MECKLENBURG COUNTY, NORTH CAROLINA 7 (2019), https://www.mdrc.org/sites/default/files/PSA_Mecklenburg_Brief1.pdf.

[77] There is already substantial scholarship on that point—for example, Eaglin, *Constructing*, *supra* note 3; Mayson, *Dangerous*, *supra* note 5, at 509.

current iteration of pretrial algorithmic governance entrenches the input problem experienced by members from low-income racially marginalized communities.

This Article uses the terms "pretrial algorithm" and "risk assessment tool" to refer to an assessment that employs statistical methods and big data to forecast the likelihood that a defendant will engage in misconduct, thereby meriting detention.[78] Both terms refer to non-automated and automated tools, which were created by a statistically derived process. Examples include the Wisconsin 2009 Risk Assessment Instrument (a checklist) and more sophisticated tools, such as COMPAS and PSA.[79]

B. *Features of Currently Employed Algorithms*

Though a relatively new phenomenon in the pretrial process, risk assessment tools have a long and controversial history in penal structures. Their initial use was in parole, where the risk scores they produced shaped parole officials' determinations of an offender's candidacy for supervised release.[80] Starting in the 2000s, whilst facing the economic consequences of mass incarceration, states and the federal government turned to sentencing risk assessment tools as a low-cost solution to reducing the incarcerated population.[81] As Jessica Eaglin explains, these tools were "meant to limit and shape the exercise of criminal law actors' discretion at the systemic level" in order to promote the release of "low risk" offenders.[82] The theory was that the risk scores produced would encourage judges to identify offenders classified as having a low risk of recidivism and to divert them to alternative programs. Despite not substantially reducing the incarcerated population in practice, risk assessment tools became and remain a popular bipartisan tool.[83]

Pretrial algorithms share the same methodology as their parole and sentencing counterparts. They predict a defendant's likelihood of pretrial misconduct by identifying the extent to which the defendant's traits correspond with the traits of other defendants who have engaged in

---

[78] I am adopting the definition provided in Mayson, *Bias*, *supra* note 3, at 2228. It is worth noting that none of the pretrial algorithms in use are machine learning.

[79] Sarah L. Desmarais & Jay P. Singh, CSG Just. Ctr., Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States 42 (2013), https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf; Northpointe, Practitioners Guide to COMPAS (2012), http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf; *Risk Assessment Factsheet: Public Safety Assessment (PSA)*, Stan. L. Sch. Pol'y Lab (May 10, 2019) [hereinafter *PSA Risk Assessment Factsheet*], https://www-cdn.law.stanford.edu/wp-content/uploads/2019/05/PSA-Sheet-CC-Final-5.10-CC-Upload.pdf.

[80] Richard A. Berk & Justin Bleich, *Statistical Procedures for Forecasting Criminal Behavior: A Comparative Assessment*, 12 Criminology & Pub. Pol'y 513, 513–16 (2013).

[81] Eaglin, *Technologically*, *supra* note 16, at 489–90.

[82] *Id.* at 504.

[83] *Id.* at 485–86.

non-appearance or arrest for alleged commission of crime.[84] These features are referred to as risk factors, which are either static or dynamic factors.[85] Static factors are unalterable traits that a defendant possesses, such as current age.[86] Dynamic factors refer to traits that a defendant currently possesses, but which tailored interventions could alter—examples include drug dependency or employment status.[87] To create these algorithms, statisticians feed a mathematical model with data about a set of pretrial defendants to identify the traits that statistically correlate with pretrial flight and crime. Since statisticians have imperfect data about pretrial flight and crime, they program the mathematical model to identify traits closely associated with a defendant's failure to appear at trial and rearrests for a new crime.[88] These traits are then deemed risk factors that developers rely on to create the pretrial algorithm. Within the algorithm, each risk factor is assigned "a number of points corresponding to how closely it is correlated with the bad outcome [of nonappearance or rearrest] in the group data."[89] The most common risk factors utilized are employment status, charges currently pending, prior custodial sentences, past record for failure to appear, drug use, residential ties, and age.[90]

Though some pretrial algorithms are made in-house, many jurisdictions that have made the transition have chosen to rely on privately developed and owned algorithms. Two popular, privately developed algorithms are COMPAS (Correctional Offender Management Profiling for Alternative Sanctions Pretrial Release Risk Scale II) and PSA (Public Safety Assessment). For this reason, the following section provides a detailed view of these two popular off-the-shelf algorithms.

1. *Correctional Offender Management Profiling for Alternative Sanctions Pretrial Release Risk Scale II (COMPAS)*

COMPAS[91] is a proprietary and commercially available pretrial risk assessment algorithm that was developed by Northpointe (now Equivant), a for-profit company. The model is the product of an analysis performed on a dataset consisting of 2,831 felony defendants on pretrial release in Kent County, Michigan, over a three-year period.[92] From that analysis, its developers created an automated algorithm consisting of eight factors: a

---

[84] Mayson, *Dangerous*, *supra* note 5, at 509.

[85] Eaglin, *Technologically*, *supra* note 16, at 490.

[86] *Id.* at 491.

[87] *Id.*

[88] Mayson, *Dangerous*, *supra* note 5, at 509.

[89] *Id.*

[90] *Id.* at 512.

[91] The COMPAS system described in this section is the 2019 version of the COMPAS system. Information about the 2012 version can be found at: NORTHPOINTE, *supra* note 79.

[92] *Risk Assessment Factsheet: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) Pretrial Release Risk Scale II (PRRS-II)*, STAN. L. SCH. POL'Y LAB (June 20, 2019) [hereinafter *COMPAS Risk Assessment Factsheet*], https://www-cdn.law.stanford.edu/wp-content/uploads/2019/06/COMPAS-PRRS-II-Factsheet-Final-6.20.pdf.

defendant's employment status; charges currently pending; prior custodial sentences; past record for failure to appear; arrests and/or criminal charges on pretrial release; drug use; most serious criminal offense; and residential ties.[93] Race is not a factor. The information needed to score the defendant is obtained by an interview and reference to court records as well as case file information.[94] The model's output is two risk scores, both in the range of 1-10.[95] One of the scores predicts the risk of failure to appear while the other forecasts the risk of a felony arrest.[96] The defendant is then assigned a risk classification. Beyond the risk factors utilized, little is publicly known about COMPAS. For instance, the weight assigned to each risk factor is neither publicly available nor provided to bail judges, prosecutors, defense counsel, or even the affected defendant.[97] Northpointe has been successful in maintaining its secrecy by enforcing it in consumer contracts with procuring jurisdictions and by asserting trade secret privilege.[98] As a result, only pretrial agencies and other licensed individuals are privy to the full inner workings of COMPAS's scoring process. Currently, two counties in California, eight counties in Wisconsin, and one county in South Carolina have adopted the algorithm.[99]

2. *Public Safety Assessment (PSA)*

PSA is a pretrial risk assessment algorithm developed by the Laura and John Arnold Foundation (now Arnold Ventures), a nonprofit foundation, which offers the model for free. The model was developed by examining a dataset consisting of 750,000 defendants on pretrial release from approximately 300 different jurisdictions across the United States over a ten-year period.[100] Like COMPAS, PSA does not use race or racial information as an input. Yet, unlike COMPAS, PSA does not utilize socioeconomic factors in its process. It only considers the following static

---

[93] *Id.*

[94] *Id.*

[95] *Id.*

[96] *Id.*

[97] *Id.*

[98] *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 772 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017) (approving the use of risk assessment tools in the sentencing context). However, it is important to note that there has been some recent efforts to limit or eliminate the ability of private companies to rely on trade secret privilege to shield the methodology of such algorithms in criminal cases. *See, e.g.*, H.B. 118, 65th Leg., 1st Reg. Sess. (Idaho 2019) (indicating that: "No builder or user of a pretrial risk assessment algorithm may assert trade secret or other protections in order to quash discovery in a criminal matter by a party to a criminal case."); Press Release, U.S. Congressman Mark Takano of California's 41st Dist., Rep. Takano Introduces the Justice in Forensic Algorithms Act to Protect Defendants' Due Process Rights in the Criminal Justice System (Sept. 17, 2019), https://takano.house.gov/newsroom/press-releas es/rep-takano-introduces-the-justice-in-forensic-algorithms-act-to-protect-defendants-due-process-right s-in-the-criminal-justice-system (noting that the Justice in Forensic Algorithms Act of 2019 was introduced to limit trade secret privilege so as to enable defendants to obtain source code of forensic algorithms).

[99] *COMPAS Risk Assessment Factsheet*, *supra* note 92; MAPPING PRETRIAL INJUSTICE, https://pretrialrisk.com/ (last visited Feb. 18, 2021).

[100] *PSA Risk Assessment Factsheet*, *supra* note 79.

factors: age at current arrest; current violent offense; charges pending at the time of the alleged offense; prior misdemeanor convictions; prior felony convictions; prior violent convictions; prior failures to appear; and prior incarceration sentences.[101] Information required to score the algorithm does not require an interview. From there, the model produces three different risk scores, each assigned a weight: failure to appear risk (scored on a scale of 0-7), new violent crime activity risk (scored on a scale of 0-13), and new criminal activity risk (scored on a scale of 0-7).[102] Weight assignments are publicly available. From that score, the defendant is provided a risk classification. PSA has been adopted statewide in Kentucky, Arizona, New Jersey, and at least ten counties outside of those three states.[103]

II. THE INPUT PROBLEM

This Part discusses the effect of the input problem on members from historically and racially marginalized communities, specifically how the input problem reproduces and entrenches the exclusion of these communities within pretrial governance and society more broadly. Its aim is not to suggest that the inclusion of these communities in algorithmic governance would necessarily lead to pretrial algorithms that would reduce or eliminate the problem of racialized pretrial incarceration. As discussed in Part III, shifting power over pretrial algorithmic governance to the most impacted communities promises no particular outcome.[104] For this reason, the aim of this Part is solely to lay out the consequences that the current iteration of pretrial algorithmic governance has had on these communities—consequences which these communities are unable to redress due to their exclusion from this form of governance.

The input problem produces multiple layers of democratic exclusion[105] for members from low-income, racially marginalized communities. The first layer concerns the tools themselves and how members from these communities are excluded from participating in the construction of algorithms used in pretrial algorithmic governance.[106] This results in pretrial algorithms that are constructed with the normative assumptions of their developers, assumptions that do not necessarily reflect the policies or outcomes that are sought by the most impacted communities.[107] The second layer relates to how these communities' exclusion from pretrial algorithmic governance facilitates their systemic marginalization in the creation and

---

[101] *Public Safety Assessment: How It Works*, ADVANCING PRETRIAL POL'Y & RSCH. 1 (2020), https://advancingpretrial.org/psa/factors/#psa-factors.

[102] *PSA Risk Assessment Factsheet*, *supra* note 79.

[103] *Id.*

[104] Simonson, *Police Reform*, *supra* note 23, at 789.

[105] Simonson, *Democratizing Criminal Justice*, *supra* note 17, at 1610–12 (discussing the multiple layers of democratic exclusion in the criminal legal system as it pertains to racialized minorities).

[106] Eaglin, *Constructing*, *supra* note 3, at 88–94.

[107] *Id.*

implementation of pretrial law, policy, and practice more broadly.[108] This results in the creation and implementation of pretrial algorithms that perpetrate the same features of the current pretrial system; features that many in these communities seek to challenge and to resist. The final layer concerns how the current iteration of pretrial algorithmic governance, through its perpetuation of the first two layers of democratic exclusion, operates to undermine efforts to contest the individual and collective political, economic, and social costs that mass pretrial incarceration, alongside other carceral practices, has had on these communities.[109] This Part will address each layer of democratic exclusion separately, since each produces a particularized and severable harm to these communities that, in combination, renders the pretrial system a site for the reproduction of racial and class stratification.

A. *Exclusion from Algorithmic Construction*

A major consequence of communal exclusion from algorithmic construction relates to the inputs relied on in currently employed pretrial algorithmic systems. These inputs operate to reproduce and maintain racialized pretrial incarceration by unfairly inflating the risk scores of racially marginalized defendants. To illustrate this problem, this Article uses two examples: (1) the prevalence of racially disparate inputs, and (2) the exclusive use of carceral risk inputs.

1. *Racially Disparate Inputs*

Nearly all pretrial algorithms utilize racially disparate inputs. Even though no pretrial algorithm explicitly uses race or racial information as an input, they tend to use inputs that are closely correlated with race. For instance, most pretrial algorithms include arrests and/or criminal records as inputs for calculating risk, both of which are systematically connected to historical and current racial inequities, including over-policing.[110] The COMPAS system considers a defendant's criminal record, past arrest record, past sentences to incarceration, and charges pending when calculating a defendant's risk of rearrest.[111] Locally created and validated algorithms,

---

[108] Simonson, *Democratizing Criminal Justice*, *supra* note 17, at 1610–12 (discussing the impact of the most impacted communities' exclusion in criminal law governance generally).

[109] Roberts, *Democratizing*, *supra* note 22, at 1605 (discussing how carceral practices on black communities "currently exclude[] [these] residents from democratic participation [and the] freedom to develop their own democratic alternatives for addressing social harms").

[110] *See generally* Anna Roberts, *Arrests as Guilt*, 70 ALA. L. REV. 987, 1020 (2019) [hereinafter Roberts, *Arrests*] (noting how arrests disproportionately visited upon racially marginalized people and do not correspond with guilt); Anna Roberts, *Convictions as Guilt*, 88 FORDHAM L. REV. 2501 (2020) [hereinafter Roberts, *Convictions*] (same).

[111] *Risk Assessment Factsheet: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) Pretrial Release Risk Scale II (PRRS-II)*, STAN. L. SCH. POL'Y LAB (June 20, 2019) [hereinafter *COMPAS Risk Assessment Factsheet*], https://www-cdn.law.stanford.edu/wp-content/uploads/2019/06/COMPAS-PRRS-II-Factsheet-Final-6.20.pdf.

such as the Colorado Pretrial Assessment Tool (CPAT), tend to take into account past arrests and criminal records.[112] Even though PSA does not explicitly consider a defendant's past arrests, past convictions and charges pending are utilized to determine a defendant's risk of failure to appear.[113]

The use of charges pending, arrests, and past convictions by these algorithms operates to unfairly inflate the risk scores of racially marginalized defendants. When used against Black defendants, these factors are more reflections of historic racial inequities rather than predictions of a defendant's propensity for future crime.[114] An example of this problem concerns arrest. Black people experience a higher rate of arrest compared to their white counterparts.[115] Though the true rate at which most criminal offenses occur among different demographic groups is unknown, studies indicate that the rate of Black arrests does not correlate to the rate of criminal offenses committed by Black people. An example of this distortion concerns drug crimes: Black people are arrested more often for drug crimes, despite committing these crimes at similar rates to white people.[116] Beyond arrest rates, past criminal convictions and sentences to incarceration are also unreliable proxies for criminal offending across different subpopulations, since racially marginalized people are also more likely to be charged, convicted, and sentenced to incarceration than their white counterparts.[117] The data also suggests that prosecutors more often pursue charges against

---

[112] COLO. ASS'N OF PRETRIAL SERVS., THE COLORADO PRETRIAL ASSESSMENT TOOL (CPAT): ADMINISTRATION, SCORING, AND REPORTING MANUAL VERSION 2, 3 (2015), https://university.pretrial .org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=47e978bb-3945-9591-7a4f -77755959c5f5.

[113] [PUBLIC SAFETY ASSESSMENT: HOW IT WORKS, *supra* note 101, at 1.

[114] Moreover, this state of affairs also inflates the myth of Black criminality which, as India Thusi has argued, "[t]rigger[s] the perception of Blacks as criminal threats to the community in the absence of harsh police tactics[,] [and] undermines arguments to recognize Black humanity." India Thusi, *Blue Lives & the Permanence of Racism*, 105 CORNELL L. REV. ONLINE 14, 23 (2020).

[115] The Department of Justice has released reports on the high rates of Black arrests in New Orleans, Ferguson, and Baltimore. *See* U.S. DEP'T OF JUST., C.R. DIV., INVESTIGATION OF THE BALTIMORE CITY POLICE DEPARTMENT 55–61 (2016), https://www.justice.gov/crt/file/883296/download; U.S. DEP'T OF JUST., C.R. DIV., INVESTIGATION OF THE FERGUSON POLICE DEPARTMENT 62–69 (2015), https://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_ department_report.pdf; U.S. DEP'T OF JUST., C.R. DIV., INVESTIGATION OF THE NEW ORLEANS POLICE DEPARTMENT 38–39 (2011), https://www.justice.gov/sites/default/files/crt/legacy/2011/03/17/nopd_rep ort.pdf; *see also* Megan Stevenson & Sandra Mayson, *The Scale of Misdemeanor Justice*, 98 B.U. L. REV. 731, 758–59 (2018) [hereinafter Stevenson & Mayson, *Misdemeanor Justice*] (discussing the disproportionate rate of Black arrests in the context of misdemeanor offenses).

[116] *See* Shima Baradaran, *Race, Prediction, and Discretion*, 81 GEO. WASH. L. REV. 157, 189–90 (2013) (finding that "[B]lack defendants are more often arrested for drug crimes even though all races commit drug crimes equally").

[117] Paul Butler, *Race and Adjudication*, *in* 3 REFORMING CRIMINAL JUSTICE: PRETRIAL AND TRIAL PROCESSES, *supra* note 42, at 211, 211–12 ("People of color are more likely to be charged with serious offenses, jailed prior to trial, convicted, and to receive a harsher sentence. These disparities exist even when factors like the severity of the crime and the criminal history of the accused person are the same.").

and offer less favorable plea deals to Black defendants in comparison to white defendants.[118]

These racial inequities dilute the predictive accuracy of the risk scores produced by present-day algorithms. The 2016 ProPublica-Northpointe debate is illustrative. In its study on the use of COMPAS in the bail hearings of 7,000 defendants in Broward County, Florida, ProPublica compared the risk classification that the algorithm assigned each defendant with their actual commission of pretrial crime within the two years following their bail hearing.[119] ProPublica researchers concluded that COMPAS was racially biased after finding that it erroneously flagged Black defendants as at high risk for pretrial crime more often than it flagged white defendants, who were correspondingly mistakenly flagged as at low risk for pretrial crime compared to Black defendants.[120] The source of the racial disparity in COMPAS was its consideration of past arrests, convictions, and sentences to incarceration in the determination of a defendant's risk score.[121] Because the Black defendants were arrested, convicted, and sentenced to incarceration more often than the white defendants in the study, the algorithmic system assigned them a high-risk classification at disproportionate rates. This produced a racial disparity even though COMPAS did not explicitly take racial information into account in its risk calculation. Importantly, COMPAS is not alone. Since most risk assessment algorithms use these factors, these systems unfairly and disproportionally falsely identify Black defendants as at high risk for pretrial crime in comparison to white defendants.

---

[118] *See, e.g.*, Besiki L. Kutateladze, Nancy R. Andiloro, Brian D. Johnson & Cassia C. Spohn, *Cumulative Disadvantage: Examining Racial and Ethnic Disparity in Prosecution and Sentencing*, 52 CRIMINOLOGY 514, 518, 527–37 (2014) (identifying racial disadvantage in plea deals offered in New York to Black defendants in comparison to other racial groups).

[119] Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[120] It is worth noting that Northpointe disputed ProPublica's findings, insisting that their algorithm was racially neutral. In support of their position, Northpointe clung to the predictive accuracy of the high-risk scores assigned, which showed that two defendants with the same base rate of arrests received a high-risk classification and were arrested for committing a pretrial crime at nearly identical rates. Both ProPublica and Northpointe's assertions were correct but premised on different metrics of algorithmic fairness. ProPublica's study was concerned with COMPAS's non-compliance with the metric of statistical parity, which led the system to falsely identify Black defendants as at high risk of pretrial crime at nearly twice the rate of white defendants. Northpointe defended itself by emphasizing COMPAS's compliance with the metric of predictive parity. Predictive parity in this context refers to the fact that two defendants labelled with a COMPAS high-risk classification committed similar rates of pretrial crime. Recent studies have demonstrated the impossibility of achieving both statistical parity and predictive parity in an algorithmic system, resulting in most pretrial algorithms complying only with the metric of predictive parity. *See* WILLIAM DIETERICH, CHRISTINA MENDOZA & TIM BRENNAN, NORTHPOINTE, COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY 9–11 (2016), https://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf (explaining Northpointe's response to the ProPublica report). *See also* Mayson, *Bias*, *supra* note 3, at 2233–36 (describing the Northpointe and ProPublica debate); Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 816–20 (same).

[121] Mayson, *supra* note 3, at 2234.

2. *Exclusive Use of Carceral Inputs*

Another input selection issue involves what this Article terms the exclusive utilization of "carceral inputs," which refers to inputs that ostensibly correspond with the risks that a defendant's release poses to public safety.[122] Nearly all currently employed algorithms are constructed to only account for the presence or absence of risk factors that correlate with a defendant's nonappearance at trial or arrest for pretrial crime. These algorithms are not designed to factor in the harms associated with pretrial detention.

The exclusive reliance on carceral inputs results in these systems obscuring and ignoring the harms associated with a defendant's pretrial detention. This is a problem because, as Crystal Yang has noted, there are individual, familial, and communal harms associated with pretrial detention.[123] Individually, pretrial detention can be a physically, emotionally, and mentally traumatizing event.[124] This traumatization can induce defendants to plead guilty despite innocence.[125] On the familial side, pretrial incarceration could lead to parental separation or even the loss of parental rights as the children of incarcerated pretrial detainees are transferred into the foster care system.[126] For the community that the defendant is part of,[127] pretrial detention endangers that community's safety by destabilizing that defendant and reducing their prospect for financial and social reintegration following their detention.[128] The sole inclusion of carceral inputs operates to maintain the negative externalities of incarceration on these low-income, racially marginalized communities. By disregarding the private and social costs that pretrial incarceration inflicts, the algorithmic construction process has led to the creation of carceral pretrial algorithms that are more likely to harm rather than to facilitate the safety of these communities.

---

[122] Shima Baradaran Baughman, *Costs of Pretrial Detention*, 97 B.U. L. REV. 1, 5–6 (2017) (stating that "[d]etainees are often victims of humiliation, rape, and other violent acts while incarcerated, and they also suffer added anxiety, stress, and a lower quality of life as a result").

[123] Yang, *supra* note 2, at 1417–28.

[124] Baughman, *supra* note 122, at 5–6.

[125] Stevenson & Mayson, *Pretrial Detention*, *supra* note 42, at 22 ("The increase in convictions [following pretrial detention] is primarily an increase in guilty pleas among defendants who otherwise would have had their charges dropped. The plea-inducing effect of detention undermines the legitimacy of the criminal justice system itself—especially if some of those convicted are innocent.").

[126] Dorothy E. Roberts, *Prison, Foster Care, and the Systemic Punishment of Black Mothers*, 59 UCLA L. REV. 1474, 1482 (2012) [hereinafter Roberts, *Black Mothers*].

[127] Jocelyn Simonson, *The Place of "The People" in Criminal Procedure*, 119 COLUM. L. REV. 249, 250–60 (2019) [hereinafter Simonson, *Place of "The People"*] (discussing how the criminal legal system promotes the false notion that the defendant's interest in pretrial release is in opposition to the community's interest in public safety).

[128] Ifeoma Ajunwa & Angela Onwuachi-Willig, *Combating Discrimination Against the Formerly Incarcerated in the Labor Market*, 112 NW. U. L. REV. 1385, 1394–98 (2018) (discussing the barriers that conviction impose on formerly incarcerated people in the labor market).

3. *Failure to Utilize Communal Knowledge*

The fact that the algorithmic construction process can produce algorithmic systems that both ignore and reify racial inequities in the pretrial process is a troubling paradox;[129] these algorithms reproduce the inequities that their employment is intended to counteract. On the one hand, the problem flows from the nature of risk assessment itself. As Sandra Mayson explains, risk assessment operates on the theory that the past will replicate itself in the future.[130] If the past is racially disparate, then the future predictions of risk by pretrial algorithms will also be racially disparate.[131]

On the other hand, the problem hinges on the failure to utilize communal information sources about the relationship between risk, crime, and the attributes of crime-committers in the algorithmic construction process. To date, developers have tended to exclusively rely on the data produced by criminal legal institutions to inform decisions about which factors should be included in algorithmic systems.[132] Yet, data from criminal legal institutions are notoriously incomplete, since most crimes are not reported[133] and crime statistics are infected with racially biased policing practices.[134] The exclusive reliance on the data produced by criminal legal institutions has facilitated the prevalence of arrests and convictions as algorithmic factors, even though these factors are not reliable proxies for a defendant's dangerousness, particularly as it pertains to members from racially marginalized communities.[135] This is because, for those communities, arrests and convictions are more likely to be wrongfully produced, and they do not necessarily correspond with factual guilt or dangerousness.[136] The net effect is that pretrial algorithms do not benefit from communal knowledge[137] and instead are constructed with racially disparate and carceral information that resultingly renders these systems less capable of predicting racially marginalized defendants' pretrial risk. The consequence is that the pretrial algorithms of today consist of carceral and racially disparate inputs that operate to reproduce existing racial disparities that serve largely to reflect

---

[129] Ajunwa, *supra* note 71, at 1673.

[130] Mayson, *Bias*, *supra* note 3, at 2224–34.

[131] *Id.* at 2251 ("Any form of prediction that relies on data about the past will produce racial disparity if the past data shows the event that we aspire to predict—the target variable—occurring with unequal frequency across racial groups.").

[132] I discuss this issue in Ngozi Okidegbe, *Discredited Data*, 107 CORNELL L. REV. (forthcoming 2022).

[133] Andrew G. Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1146 (2017).

[134] Mayson, *Bias*, *supra* note 3, at 2264.

[135] Berk & Bleich, *supra* note 81, at 538–40 (finding that the use of past misdemeanor arrests and the age of an offender's earliest arrest were not reliable predictors of future offending for Black juvenile defendants as compared to white juvenile defendants).

[136] *See, e.g.*, Roberts, *Arrests*, *supra* note 110, at 1020 (noting how arrest do not connote factual guilt); Roberts, *Convictions*, *supra* note 110, at 2509 (noting how convictions can be wrongfully produced).

[137] Okidegbe, *supra* note 134.

the racial inequities existing in the dispensing of arrests and convictions within the criminal legal system.

B. *Exclusion from Pretrial Governance*

Communal exclusion from pretrial algorithmic governance operates to reproduce these groups' marginalization in pretrial governance. It is true that before the deployment of pretrial algorithms, these communities were democratically excluded from full participation in the creation, implementation, and oversight of pretrial law, policy, and practices. As Jocelyn Simonson's work has demonstrated, this democratic exclusion results from their political disempowerment.[138] Unlike their wealthier and whiter counterparts, members from these historically marginalized communities have diminished electoral power to vote for bail judges and legislators who reflect their communal preferences around pretrial justice.[139] For this reason, pretrial legislation, policy, and practice are responsive to the preferences of socioeconomically powerful citizens, who have the financial resources to influence the appointment and direct election of bail judges who share their background and views around the pretrial system. More importantly, these financial resources enable powerful groups to build, as Sabeel Rahman notes, an "ecosystem of lobbying, advocacy, and model legislation"[140] that impacts pretrial law, policy, and practice. For the above reasons, the most impacted communities have diminished political power to render those in charge of pretrial governance accountable to their needs and interests.

One effect of these communities' political powerlessness relates to their over-incarceration in the pretrial system. As Samuel Wiseman's work has demonstrated, the current political and social climate incentivizes bail judges to over-detain defendants awaiting trial[141]—the reason being that bail judges experience minimal consequences for detaining a low-risk defendant, yet face intense public backlash for the release of a defendant who ultimately commits another crime or flees the jurisdiction.[142] For bail judges, the cost of a release that is perceived as erroneous might mean the loss of their judgeship.[143] In contrast, no individual judge faces a real prospect of impeachment or loss of an election from over-incarceration practices, since members from those communities lack the political power to render these judges democratically accountable to them. The state of affairs has meant that decisions around pretrial release, detention, and surveillance are rarely

---

[138] Simonson, *Democratizing Criminal Justice*, *supra* note 17, at 1611–12,1621.

[139] Dorothy E. Roberts, *Foreword: Abolition Constitutionalism*, 133 HARV. L. REV. 1, 37 (2019) (discussing how the collateral penalties from criminal punishment have diluted Black political power).

[140] K. Sabeel Rahman, *Policymaking as Power-Building*, 27 S. CAL. INTERDISC. L.J. 315, 318 (2018) (discussing this problem in the administrative context).

[141] Wiseman, *supra* note 55, at 431.

[142] For more information on this point, see generally RACHEL ELISE BARKOW, PRISONERS OF POLITICS: BREAKING THE CYCLE OF MASS INCARCERATION (2019).

[143] Wiseman, *supra* note 55, at 422.

in line with the very communities that stand to be harmed by large-scale pretrial incarceration.[144]

Though pretrial algorithms did not inaugurate the democratic exclusion that these communities traditionally face within pretrial governance, their utilization operates to reinforce it. This is because communities, due to their exclusion from algorithmic governance, are unable either to affect the algorithm's construction or to viably challenge its results in a pretrial hearing. To understand this problem, it is important to consider bottom-up and confrontational practices[145] that have emerged in recent years to contest over-incarceration in the pretrial context. One important practice has been the rise of community bail funds, which contest over-incarceration practices within the pretrial system by systematically bailing out low-income defendants, whose release enhances community safety.[146] As Jocelyn Simonson has noted, these communal bail funds reflect and inject a communal voice about pretrial release practices.[147] By bailing out a defendant, a community bail fund shows that at least one segment of the community contests the bail judge's assessment that the defendant's safe release must hinge on the payment of an unaffordable bail amount. These bottom-up practices have been an important disruptive practice that has rendered visible the disconnect between the pretrial system's operation and the views and values of the most impacted communities.[148]

Despite their efficacy, such bottom-up practices are ineffective at contesting and impacting pretrial interventions that are shaped by pretrial algorithms. This is because many jurisdictions that utilize pretrial algorithms have also eliminated the practice of cash bail, conditioning decisions around release, detention, and surveillance instead on a defendant's perceived

---

[144] Simonson, *Democratizing Criminal Justice*, *supra* note 17 (noting how criminal law, policy, practice, and decision-making is rarely responsive to those most harmed by it).

[145] Amna A. Akbar, *Law's Exposure: The Movement and the Legal Academy*, 65 J. LEGAL EDUC. 352, 364 (2015) ("[Social movements] make[] clear that the ordinary channels of accountability cannot be relied on. . . . Contrary to traditional litigation and voting disconnected from a larger campaign, the movement's approaches are confrontational with the state, or turn away together to come back another day stronger.").

[146] There are a number of community bail organizations actively bailing out low-income defendants, such as One Community, NorCal Resist, and Chicago Community Bond Fund. *Ways to Support the Black Community and Combat Racism*, ONE CMTY., https://www.onecommunity.co/ways-to-support-the-black-community-and-combat-racism (last visited Feb. 25, 2021); *NorCal Resist Activist Bail & ICE Bond Fund,* NORCAL RESIST, https://actionnetwork.org/fundraising/ncrbailfund/ (last visited Feb. 25, 2021); *Our Mission*, CHI. CMTY. BOND FUND, https://www.chicagobond.org/ (last visited Feb. 25, 2021).

[147] *See* Simonson, *Nullification*, *supra* note 54, at 591 ("Community bail funds provide to the public real-life examples of indigent defendants returning to court without having undermined public safety, despite an expert judicial determination that personal money was needed to prevent flight and mayhem.").

[148] It is important to note that these practices are disruptive, but they also can unintentionally operate to legitimate the system. *Id.* at 631. ("On the one hand, one might object to community bail funds—especially when they resemble bail nullification—as a subversion of the rule of law; and on the other hand, one might worry that a belief in the power of community bail funds risks legitimizing an unfair procedural scheme.").

riskiness.[149] Without the mechanism of cash bail, decisions around pretrial detention cannot be influenced or counteracted by communal practices. Community bail funds, for example, are powerless to facilitate the release of a defendant, whose detention is partially conditioned on a risk prediction, and not on cash bail. To be clear, the aim of recognizing this problem is not to suggest a preference for cash bail. Cash bail is a socioeconomic and racially disparate practice that has itself been a major contributor to the crisis of racialized pretrial incarceration.[150] Rather, the point is that pretrial algorithmic governance means the foreclosure of avenues for disruptive and influential communal participation in pretrial governance. This consequence has even pushed a few community groups to seek inclusion in pretrial algorithmic governance, despite their repudiation of the algorithm project.[151] Given this, these communities' exclusion from pretrial algorithmic governance serves to reinforce their marginalization in the crafting and implementation of the pretrial laws and policies that have promoted their over-incarceration—therein hampering current efforts to reform or to dismantle and rebuild the pretrial system to be in line with communal values and viewpoints. [152]

C.   *Exclusion from Full Participation in a Democratic Society*

Both communal exclusion from algorithmic construction and this exclusion's reinforcement of these groups' marginalization within pretrial governance produce a system primed to uphold the racial status quo with no effective avenue for communal resistance. This facilitation of the status quo leads to a third layer of the input problem, which is pretrial algorithmic governance's exacerbation of the democratic harms that the criminal legal system already enacts on marginalized communities. A number of scholars, such as Dorothy Roberts, Janet Moore, Jocelyn Simonson, and others, have argued that the criminal legal system enacts democratic harms on racially marginalized communities by diluting their political, economic, and social power to the point of denying them full participation in our democracy.[153] By its exclusion of these communities, pretrial algorithmic governance, alongside other criminal laws and practices, functions as a site that

---

[149] Diana Dabruzzo, *New Jersey Set Out to Reform Its Cash Bail System. Now, the Results Are In.*, ARNOLD VENTURES (Nov. 14, 2019), https://www.arnoldventures.org/stories/new-jersey-set-out-to-reform-its-cash-bail-system-now-the-results-are-in/.

[150] Arnold et al., *supra* note 64, at 1917.

[151] *See* LEADERSHIP CONF. EDUC. FUND, *supra* note 9, at 8 (discussing lack of community consultation around algorithmic construction).

[152] *See infra* Section II.C.

[153] Janet Moore, *Democracy Enhancement in Criminal Law and Procedure*, 2014 UTAH L. REV. 543, 547 (2014); Roberts, *Democratizing*, *supra* note 22, at 1605; Simonson, *Police Reform*, *supra* note 23, at 806–07; Monica C. Bell, *Anti-Segregation Policing*, 95 N.Y.U. L. REV. 650, 734 (2020); *see also* Zina Makar, Detention, Disenfranchisement, and Doctrinal Integration (Oct. 13, 2020) (unpublished manuscript) (on file with author) (exploring how pretrial detention de facto strips detainees of their right to vote).

reproduces and contributes to the democratic harms experienced by these communities, thus facilitating a racialized system.[154]

On this basis, the input problem sets pretrial algorithmic governance up to reproduce the pretrial system's political, economic, and social costs on the most impacted communities. This is concerning since these costs disentitle these communities from exercising their rights to full citizenship.[155] The pretrial system of detention adversely affects the democratic participation of these communities. On account of the plea-inducing impact of pretrial incarceration, many exit the system with a criminal record that temporarily or permanently restricts their right to vote.[156] These restrictions dilute these communities' political power to orient the pretrial system away from policies that promote the over-incarceration of their community members.[157] Economically, pretrial incarceration strains wealth production as financial, material, and intellectual resources are diverted to the jail system.[158] Alongside post-conviction incarceration, pretrial incarceration operates as a redistribution of wealth from the poor to the state, extracting the limited resources of low-income communities of color, which in turn disrupts intergenerational wealth transfers and perpetuates the racial cycle of poverty.[159] Socially, the pretrial system overextends social kinship networks, as those left behind are forced to expend time and limited resources assuming the financial and caregiving obligations of those who are incarcerated.[160] The combination of these consequences creates a racial, gendered, and class geography of disrupted social kinship networks, economic disenfranchisement, and political estrangement that operates to deny members from low-income Black communities full citizenship, all of which the current iteration of pretrial algorithmic governance contributes to and entrenches.

---

[154] Simonson, *Police Reform*, *supra* note 23, at 807.

[155] *See* Dorothy E. Roberts, *The Social and Moral Cost of Mass Incarceration in African American Communities*, 56 STAN. L. REV. 1271, 1277 (2004) [hereinafter Roberts, *Mass Incarceration*] (contending that incarceration generally "has become a systemic aspect of community members' family affairs, economic prospects, political engagement, social norms, and childhood expectations for the future.").

[156] *See* Ekow N. Yankah, *Good Guys and Bad Guys: Punishing Character, Equality and the Irrelevance of Moral Character to Criminal Punishment*, 25 CARDOZO L. REV. 1019, 1030–31 [hereinafter Yankah, *Good Guys*] (2004) (discussing the racial dimension of felon disenfranchisement)

[157] *Id.* at 1031.

[158] Roberts, *supra* note 155, at 1282.

[159] *Id.*

[160] *See* Roberts, *Black Mothers*, *supra* note 126, at 1482 (noting how the large-scale incarceration of Black fathers disrupts family life and places an insurmountable burden on Black women caregivers, who struggle to take upfill the financial and social void that the incarcerated person leaves behind). Ekow N. Yankah, *Punishing Them All: How Criminal Justice Should Account for Mass Incarceration*, 97 RES PHILOSOPHICA 185, 190 (2020) [hereinafter Yankah, *Punishing Them All*] ("The incarceration of large numbers from particular communities undermines the fabric that sustains essential community functions.").

III. Solving the Input Problem

Part II set out the particular harms that the input problem imposes on members from low-income racialized communities. In this Part, this Article discusses the most promising redress to the input problem: shifting power over pretrial algorithmic governance to members from the most impacted communities. To do this, it starts by discussing the growing consensus around the need for public input into algorithmic design. It then demonstrates the limitations associated with approaches to the input problem that do not shift power over algorithmic governance to these communities. Using the example of Pennsylvania, which held a series of public hearings around its decision to employ a sentencing algorithm, it contends that such approaches, though politically palatable and implementable, largely maintain the input problem whilst presenting the risk that communal participation will be misconstrued as communal approval. It then sets out the dimensions of this Article's power-shifting model. The promise and anticipated objections to this power-shifting model are discussed in detail in Part IV and Part V.

A. *Consensus Around Public Participation*

Before discussing ways to resolve the input problem, it is necessary to note the growing consensus around the importance of public participation in algorithmic governance. While algorithmic governance was in its infancy, there was a contentious debate about the necessity of public input.[161] However, as algorithmic systems and structures have mushroomed in recent years, a growing consensus about the importance of public participation in algorithmic governance has emerged.[162] It has become increasingly apparent that the design and deployment of algorithmic systems could benefit from public feedback, especially since these systems are not created in a political

---

[161] Some claimed that soliciting public input might undermine the promising features of algorithmic governance, namely its objectivity and insulation from the arbitrariness and bias of human decision making. Since members of the public lack technical and policy expertise, giving credence to their viewpoint risked diluting the accuracy and efficiency of algorithmic systems whilst also delegitimizing the entire algorithmic project. Moreover, supporters of this viewpoint have tended to contest the idea that public participation in algorithmic governance is a precondition for achieving algorithmic accountability and legitimacy.

[162] New York City has developed a task force highlighting the importance of public participation in algorithmic governance. *See, e.g.*, New York City Automated Decision Sys. Task Force, New York City Automated Decision Systems Task Force Report 22 (2019), https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf (recommending the broadening of public discussions on automated decision systems that New York agencies and offices use in the course of providing public services). A few organizations have also created partnerships between communities and state actors to facilitate public participation around technology. *See Responsible Use of Policing Technology*, Policing Project N.Y.U. Sch. L., https://www.policingproject.org/policing-tech-landing (last visited Sept. 26, 2021) (discussing the importance of public input in the use of policing technologies).

or social vacuum.[163] Constructing, implementing, and overseeing any algorithmic system requires technocrats to make highly discretionary and value-laden decisions that shape how law and public policy are applied. For instance, as Jessica Eaglin has noted in the context of sentencing algorithms, decisions around risk thresholds are unavoidably normative because it is impossible to determine the cutoff between different risk thresholds without confronting an unavoidable value judgment about the level of risk that is socially acceptable.[164] These realities have substantially undermined the notion that algorithmic systems are neutral, impartial, or apolitical, or should be devoid of public input.[165] At the same time, concerns around algorithmic accountability have bolstered ongoing support for public participation in algorithmic governance, the idea being that the legitimacy of algorithmic governance hinges on its ability to engage with and be responsive to public input.[166] For adherents of this view, algorithmic governance needs to afford the same types of participatory mechanisms that have been integral to governmental processes, particularly in the administrative realm.

B. *Limits of Non-Power-Shifting Approaches*

As consensus coalesces around the importance of public input, a few jurisdictions have sought to mediate public participation in algorithmic governance. All these approaches have transpired under a participatory model of communal involvement that revolves around the ex-post solicitation of public input from a variety of stakeholders in forms ranging from stakeholder meetings[167] to public hearings.[168] This model of communal

---

[163] Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. DAVIS L. REV. 1067, 1085–86 (2018); John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1803–05 (2018); Ngozi Okidegbe, *When They Hear Us: Race, Algorithms and the Practice of Criminal Law*, 29 KAN. J.L. & PUB. POL'Y 329, 334 (2020) (noting transparency as a key precondition for facilitating communal participation in algorithmic governance); Eaglin, *Constructing*, *supra* note 3, at 61; Mailyn Fidler, *Local Police Surveillance and the Administrative Fourth Amendment*, 36 SANTA CLARA HIGH TECH. L.J. 481, 560 (2020) (advocating for local input in policing technologies).

[164] Eaglin, *Constructing*, *supra* note 3, 73.

[165] *See* NOBLE, *supra* note 31, at 2 (discussing the problem that algorithmic systems are not neutral and are designed with the sexist, racist, and classist ideas of their designers, in the context of algorithms designed in the search engine context); Sarah Valentine, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control*, 46 FORDHAM URB. L.J. 364, 379 (2019) (advocating for public input in algorithmic governance).

[166] Requiring that the value-laden decisions around algorithmic governance be subject to public debate and scrutiny may force developers and other technocrats to justify their decisions around algorithmic systems, promoting accountability. *See* David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REGUL. 800, 846 (2020).

[167] *See* PRANITA AMATYA, SHELBY KING, SHELBY MCNABB & HEIDI SCHULTHEIS, BAIL REFORM IN CALIFORNIA (2017), https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ash x?DocumentFileKey=835f283a-e9fc-9c56-28bb-073a9bcb1dbf (providing an example of a process dependent on stakeholder meetings)

[168] *See* Stephanie Wykstra, *Can Racial Bias Ever Be Removed From Criminal Justice Algorithms?*, PACIFIC STANDARD (July 12, 2018) [hereinafter Wykstra, *Racial Bias*], https://psmag.com/social-justice/removing-racial-bias-from-the-algorithm (discussing the fact that the Pennsylvania Commission on Sentencing held a public hearing around the use of a sentencing algorithm in Pennsylvania); NEW

involvement takes inspiration from participatory processes utilized in the administrative sphere. As Richard A. Bierschbach and Stephanos Bibas have explained, a cornerstone of the administrative approach to communal participation is to allow "citizens [to] communicate their information and views, but . . . the governmental decision-maker [has the] ultimate power to balance the various inputs and make a final decision."[169]

The problem with this approach is not communal exclusion. Rather, members from these communities or relevant community groups are often included as stakeholders at these sessions. The issue with this approach is that the input problem is largely maintained by design. Pennsylvania's mediation of public participation with its sentencing algorithm is illustrative. In 2010, Pennsylvania passed legislation requiring its Sentencing Commission (the "Commission") to develop a risk assessment instrument as part of its strategy to reduce the state's incarcerated population.[170] When resistance to this sentencing algorithm project gained momentum, the Commission held a series of public hearings to solicit public feedback around the algorithm's construction to quell concerns.[171] The process included various stakeholders, including community groups representing members from the most impacted communities.

The effort was hailed by many as an exemplary approach to mediating public participation in algorithmic governance.[172] However, key features of this participatory process locked in place the input problem and its ensuing consequences. First, the entire participatory process solicited ex-post input around algorithmic governance. By situating public participation after the approval of the use of algorithmic governance, the most impacted communities were stripped of the power to take up first-order questions about the nature and value of algorithmic governance itself in the sentencing context.[173] Second, the process placed exclusive control over the algorithm's inputs within the hands of the Commission.[174] This left little room for

---

YORK CITY AUTOMATED DECISION SYS. TASK FORCE, *supra* note 162, at 15–16 (contending that the task force held a number of meetings with stakeholders around the use of automated decisions in the public sector). *But see* AI NOW INST., CONFRONTING BLACK BOXES: A SHADOW REPORT OF THE NEW YORK CITY AUTOMATED DECISION SYSTEM TASK FORCE 14–16 (Rashida Richardson ed., 2019) (criticizing the small number of stakeholders consulted by the New York City Automated Decision Systems Task Force).

[169] Richard A. Bierschbach & Stephanos Bibas, *Notice-and-Comment Sentencing*, 97 MINN. L. REV. 1, 21 (2012).

[170] The instrument was intended to aid judges in identifying low-risk offenders for alternative sentencing programs. 42 PA. CONS. STAT. § 2154.7 (2020).

[171] Stephanie Wykstra, *Just How Transparent Can a Criminal Justice Algorithm Be?*, SLATE: FUTURE TENSE (July 3, 2020, 8:00 AM), https://slate.com/technology/2018/07/pennsylvania-commission-on-sentencing-is-trying-to-make-its-algorithm-transparent.html.

[172] Wykstra, *Racial Bias*, *supra* note 168.

[173] *See, e.g.*, Simonson, *Police Reform*, *supra* note 23, at 791–92 (noting how the most impacted communities are excluded from first order questions about "how the state should provide safety and security") .

[174] This problem is only compounded by the vulnerability of institutional stakeholders to capture. *See* Rachel E. Barkow, *Insulating Agencies: Avoiding Capture Through Institutional Design*, 89 TEX. L.

members from the most impacted communities to contest and to resist the use of inputs that would promote the over-incarceration of their community members and resultingly threaten their community safety. For instance, though community organizing drove the Commission to eventually abandon the use of past arrests as a factor in the calculation of a risk score,[175] the Commission did not remove other proxies, such as past convictions, that would unfairly inflate the risk scores of Black offenders.[176] Since the Commission had sole control and decision-making power over the risk assessment tool, it had discretion to determine which recommendations it would take on. This left the participatory process with no mechanism to prioritize the feedback of Black Pennsylvanians, despite the fact that their disproportionate rate of incarceration—which is approximately nine times that of their white counterparts[177]—renders them uniquely vulnerable to the sentencing algorithm's outputs. These dynamics "reinscrib[ed] rather than dismantl[ed] existing power imbalances[,]"[178] enabling the Commission to adopt a sentencing algorithm positioned to reproduce existing racial disparities.[179] Yet, disturbingly, despite the lack of communal control over the algorithm's purpose and design, the sentencing algorithm might benefit

REV. 15, 21 n.23 (2010) (describing capture as undue "responsiveness to the desires of the industry or groups being regulated").

[175] Asli Bashir, *Pennsylvania's Misguided Sentencing Risk-Assessment Reform*, REGUL. REV. (Nov. 5, 2020), https://www.theregreview.org/2020/11/05/bashir-pennsylvania-misguided-sentencing-risk-assessment-reform/. It is important to note the fact that discriminatory effects of including of past arrests as a factor in the algorithm was recognized by a few of the commissioners. For instance, Rachel Lopez problematized and advocated to remove the use of arrests as an algorithmic factor during several commission meetings. *See, e.g.*, PA. COMM'N ON SENT'G, COMMISSION MEETING MINUTES: MAR. 3, 2016 5, 7 (2016), http://pcs.la.psu.edu/about-the-commission/meetings/meeting-minutes/prior-years/2016/march-2016/view.

[176] Elizabeth Hardison, *After Nearly a Decade, Pa. Sentencing Commission Adopts Risk Assessment Tool over Objections of Critics*, PA. CAPITAL-STAR (Sept. 5, 2019), https://www.penncapital-star.com/criminal-justice/after-nearly-a-decade-pa-sentencing-commission-adopts-risk-assessment-tool-over-objections-of-critics/.

[177] *See* ASHLEY NELLIS, SENT'G PROJECT, THE COLOR OF JUSTICE: RACIAL AND ETHNIC DISPARITY IN STATE PRISONS 7 (2016), https://www.sentencingproject.org/publications/color-of-justice-racial-and-ethnic-disparity-in-state-prisons/(showing the racial disparity in rate of incarceration).

[178] Simonson, *Democratizing Criminal Justice*, *supra* note 17, at 1610.

[179] This problem was brought to the Commission's attention by a 2019 independent assessment by a Carnegie Mellon University research team. That report recommended that the sentencing algorithm's use be limited to low-risk defendants, after concluding that the sentencing algorithm only accurately identified high-risk defendants forty-eight percent of the time and was particularly inaccurate in relation to Black defendants. DAVID MITRE BECERRIL, CHRIS BELL, KATIE LEFEVRE, LAUREN LIN, WILSON MUI & KARAN SHAH, VALIDATION AND ASSESSMENT OF PENNSYLVANIA'S RISK ASSESSMENT INSTRUMENT 52–53 (2019), http://pcs.la.psu.edu/guidelines/sentence-risk-assessment-instrument/sentence-risk-assessment-research-archives/validation-and-assessment-of-pennsylvanias-risk-assessment-instrument.-heinz-college-system-synthesis-project-may-22-2019-1/view. It should also be noted that the Commission made small alterations to the sentencing algorithm in response to the Carnegie Mellon report, but rejected the recommendation to use the tool for low-risk offenders. *See* PA. COMM'N ON SENT'G, RISK ASSESSMENT UPDATE: STAFF'S RESPONSE TO CARNEGIE MELLON UNIVERSITY'S EXTERNAL REVIEW 2–3 (2019), http://pcs.la.psu.edu/guidelines/sentence-risk-assessment-instrument/sentence-risk-assessment-research-archives/commission-response-to-external-review/view

from the illusion that its construction was the product of community consultation, which, in turn, will facilitate its legitimation in the state.

The Pennsylvania experience illustrates the limits of redressing the input problem by merely facilitating greater public participation in algorithmic governance. By not facilitating the empowerment of members from the most impacted communities, such processes will inevitably reinscribe existing power imbalances that operate to undermine efforts by these communities to resist or influence algorithmic governance. This state of affairs begs the question of whether any efforts short of shifting power over algorithmic governance could correct the input problem. For the reasons discussed below, this Article suggests that while such efforts can ameliorate the first layer of the input problem by facilitating communal involvement in algorithmic construction, these efforts are unable to ameliorate the latter two layers of the input problem, since they reproduce it.

To show this, the following section considers a series of approaches that would center the participation of the most impacted communities without endowing them with decision-making power. This kind of participation would allow members from these communities to lodge their views about algorithmic governance but would not endow these communities with a veto regarding if and on what basis such governance would be pursued in the pretrial system.

1. *Focus Groups*

The first option would mean holding focus groups comprised of members from communities most impacted by incarceration. One possibility in this vein would be for private companies to hold such focus groups as a precondition in designing an algorithm for pretrial use. Procurement contracts could require such a process. Beyond the recruitment and sampling bias associated with the use of focus groups,[180] the main problem is that this approach endows private companies with the ultimate control and decision-making power over the algorithmic formula. Companies—rather than members from these communities—would retain power to determine which input to incorporate or disregard, leaving intact the two latter layers of the input problem. To compound this issue, the focus group process can operate to legitimatize the end algorithmic product under the guise of community approval, even if all or most of the communal input was disregarded. For this reason, this option would not fix the input problem.

2. *Public Hearings*

The second option would require the holding of public hearings solely designed to obtain input from the most impacted communities. This would differ from Pennsylvania's approach since it would prioritize the feedback

---

[180] *See generally* David L. Morgan, *Why Things (Sometimes) Go Wrong in Focus Groups*, 5 QUALITATIVE HEALTH RSCH. 516, 517–20 (1995) (noting the problems of focus groups).

of the most-impacted communities rather than position the most-impacted communities as merely one among many relevant stakeholders.[181] Two problems arise with this approach. One problem is the exacerbation of power imbalances that already exist in these communities. Given that those attending these hearings tend not to receive financial compensation or reimbursement, participation would likely be skewed towards the most financially stable and educated members in these communities to the detriment of the most marginalized members. Furthermore, like the focus group model, this option neither facilitates the incorporation of these communities' views into the algorithm nor would it afford them a mechanism to resist the use of the algorithm itself. Both issues mean that adopting this option would entrench the latter two layers of the input problem.

3. *Appointed Citizen Boards*

The third option would be to create appointed citizen boards where community members could serve and provide input to the ultimate designer of the algorithm. To create such a board, one might borrow from the policing context, which has seen a rise in informal policing advisory citizen boards that provide input on the practices and policies of police forces.[182] A major flaw with this approach is that there is no mechanism to contest the use of the algorithm or to ensure their communal viewpoints around the algorithm's use and purpose are taken on. As such, this approach risks regulating communal feedback to ex-post decisions around algorithmic adoption, construction, and oversight, which, like Pennsylvania's approach, strips communities of involvement in first-order questions about the algorithm's design, purpose, and value.

These shortcomings illustrate how non-power-shifting approaches maintain key consequences of the input problem. Even with these shortcomings, the benefits of the participatory approach cannot be ignored. This model is able to counteract the first layer of the input problem. One promising feature of that counteraction is that this model could enable members from oppressed communities to raise their viewpoints about algorithmic governance. This might be empowering in that it might afford a degree of accountability to these communities by incentivizing decision-makers to justify their algorithmic system's inputs and outputs.[183] This process might influence decision-makers to make improvements to the algorithmic systems used along the lines that these communities proffered,

---

[181] This dynamic risks reproducing the Pennsylvania experience. For more information, see Section II.B

[182] One example is the Community Relations Bureau's Community Response Squad that advises the Phoenix police force. *See Community Response Squad*, CITY PHOENIX, https://www.phoenix.gov/police/community-relations/community-response-squad (last visited Mar. 13, 2021). For criticisms of community police models, see Joseph Rukus, Mildred E. Warner & Xue Zhang, *Community Policing: Least Effective Where Need Is Greatest*, 64 CRIME & DELINQ. 1858, 1858 (2018).

[183] Bierschbach & Bibas, *supra* note 169, at 23 (discussing how public input in the administrative arena improves accountability "by obligating agencies to justify their actions publicly, ensuring that they are 'relatively informed and responsive to public needs'").

potentially decreasing these communities' experience of pretrial detention. Additionally, this model is politically palatable. If algorithmic governance is inevitable, it may be worth it to seek approaches that partially redress the input problem. This is especially true if more radical approaches to challenging and redressing racialized pretrial incarceration, such as full abolition, remain politically infeasible and the large-scale reduction of pretrial detention, such as provided for in the New York Bail Reform Law, does not guard against the continued use of risk assessment tools.[184] On this basis, counterintuitively, seeking at least inclusion in algorithmic governance, even if it is not control, might serve to facilitate racial justice alterations to the formulas behind the algorithms implicated in the over-incarceration of racially marginalized defendants. Given these potential benefits, it is conceded that the participatory model is preferable to the status quo.

Yet, the major downside of the participatory model is that it largely leaves intact the other ensuing consequences of the input problem, which concern pretrial governance's contribution to the democratic exclusion experienced by racially marginalized communities both within and outside of pretrial governance. The main reason for this result is that the participatory model affords no lever for facilitating the incorporation of the communal input solicited. Decision-makers are free to reject or incorporate communal feedback as they wish, leaving these marginalized communities without an avenue to reform or dismantle the pretrial algorithms that are adversely affecting their daily lives. Moreover, this model risks skewing outcomes towards the preferences of the most powerful stakeholders, because, as K. Sabeel Rahman and Jocelyn Simonson note, "some constituencies already possess greater capacity for power and influence" than others.[185] Any process for mediating public input that is not specifically designed to counteract this operation of privilege guarantees amplifying the voices of members from whiter and wealthier communities whilst muting the voices of less privileged groups. For this reason, when applied to the context of pretrial algorithmic governance, the participatory model risks marginalizing or alienating members in the most impacted communities that are opposed to all or key aspects of the algorithm project. The materialization of either outcome means the reproduction and maintenance of a large swath of the input problem. Given this, power-shifting, which is discussed in the next section, remains the most viable approach to redressing the input problem but poses a different series of concerns, which are discussed in Part V.

---

[184] New York Bail Reform Law, *supra* note 40, at 182, 184.

[185] Rahman & Simonson, *supra* note 29, at 689–90.

C. *Power-Shifting Approach*

The power-shifting approach is one that would shift power over all or at least key aspects of pretrial algorithmic governance to the most impacted communities. It is connected to a particular conceptualization of democratic participation[186] that has been theorized and advocated for in criminal law scholarship and administrative scholarship, most recently by Jocelyn Simonson and K. Sabeel Rahman.[187] Its central contention is that substantive or procedural policies alone cannot ameliorate or dismantle racial and socioeconomic inequities unless those policies attend to the power imbalance partially responsible for these inequities.[188] From this starting point, it holds that attending to power means shifting the allocation of power within institutional structures toward the most marginalized members of society, endowing them with direct power to "influence policy outcomes and control the distribution of state resources."[189] This shift in power aims to provide mechanisms to promote the democratic participation of vulnerable groups that have traditionally been denied a share of decision-making power.[190]

The focus on redistribution of power is what differentiates this model from the participatory model of communal involvement. It recognizes that inclusion in governing structures without attending to the allocation of power cannot materially reform or dismantle the status quo. Rather, such inclusion serves only to reproduce and legitimate racialized systems that maintain the subordination of oppressed communities.[191] For this reason, in contrast to the participatory approach, the power-shifting model specifically seeks to foster the participation of communities most impacted by the carceral state. In so seeking, its design aims to counteract the marginalization that these communities face within governing structures and within society by centering these communities.

1. *Potential Power-Shifting Scenarios*

This Section sets out one scenario in which a power-shifting model of communal involvement could be actualized in pretrial algorithmic governance: bail commissions at the state and federal levels consisting of

---

[186] CHANTAL MOUFFE, AGONISTICS: THINKING THE WORLD POLITICALLY 7 (2013) ("Conflict in liberal democratic societies cannot and should not be eradicated, since the specificity of pluralistic democracy is precisely the recognition and the legitimation of conflict. What liberal democratic politics requires is that the others are not seen as enemies to be destroyed, but as adversaries whose ideas might be fought, even fiercely, but whose right to defend those ideas is not to be questioned. To put it in another way, what is important is that conflict does not take the form of an 'antagonism' (struggle between enemies) but the form of 'agonism' (struggle between adversaries).").

[187] Rahman & Simonson, *supra* note 29, at 689–93; Rahman, *supra* note 140, at 318–19; .

[188] MOUFFE, *supra* note 186, at 81.

[189] Rahman & Simonson, *supra* note 29, at 692.

[190] *Id.* at 689–92.

[191] Simonson, *Police Reform*, *supra* note 23, at 807 (contending that layers of democratic exclusion in the criminal legal system have facilitated the "reproducing and legitimizing an unequal and racialized system of justice").

members from low-income, racially marginalized communities with the principal task of determining if and on what basis pretrial algorithms are used in the pretrial system. The scenario is premised on the creation of a new institutional structure, the reason being that such a structure provides the most promising avenue for effective power over pretrial algorithmic governance.

For this proposed institutional structure to have a prospect of power- shifting, two conditions must be met: (1) the endowment of decision-making power to the most impacted communities, and (2) representation of the most impacted communities on the commission.[192]

i.   Decision-Making Power

Decision-making power must be exercisable by the most impacted communities. This level of control is on a continuum, ranging from exclusive non-reviewable control to control over only key aspects of the adoption, implementation, and oversight of pretrial algorithms. At the very least, power-shifting requires communities to have a mechanism to mandate the incorporation of their viewpoints and values into algorithmic governance. Having a mechanism means having control over the algorithm. Merely allowing communal representatives to participate in or provide input on algorithmic governance is insufficient. Within the spectrum, one vision of power-shifting could be a commission consisting exclusively of members from the most impacted communities tasked with administrating, implementing, and overseeing pretrial algorithms, assuming that body decides to pursue pretrial algorithmic governance. Under that vision, technocrats—such as independent data scientists, developers, and policymakers—may or may not have a role to play on the commission. Any role that they may have would be subject to communal approval. On the other side of the spectrum, however, another vision would be a commission consisting of technocrats and community representatives with the primary task of determining if and how pretrial algorithmic governance operates in the jurisdiction. Under this vision, community representatives would only have veto power over key aspects of pretrial algorithmic governance— aspects that would have out-sized effects on their community's safety. For example, community representatives could have a veto power over the adoption of pretrial algorithmic governance in the first instance and the utilization of any input within the pretrial algorithm that would increase the use of pretrial incarceration. Other decisions around pretrial algorithmic governance could be decided by a simple majority. This latter vision would lead to a blending of communal and technocratic expertise that could facilitate the shaping of algorithmic governance to reflect communal needs and values.

---

[192] Rahman & Simonson, *supra* note 29, at 682, 723.

ii.   Composition of the Commission

Power-shifting requires a shift in who exercises decision-making power. This means that representatives on the commission must be the Black people living in the communities most affected by the use of pretrial algorithmic governance. Yet, one difficult issue in implementing this approach is defining eligibility. Who should be eligible for community representative positions? Should eligibility be defined spatially, by self-identification, or by experience? The difficulty is compounded by the fact that every person whose body is ascribed with the racial category of Black experiences the political, social, and economic consequences of pretrial incarceration, even if indirectly. However, only some will have the experience required to provide input approximating the needs of the most impacted. For this reason, only those who live in low-income communities and have direct experience with incarceration must hold these positions. This direct experience could take the form of: (1) having been formerly incarcerated; (2) being the immediate family member of a current or formerly incarcerated person; or (3) being the direct victim of a crime. Three reasons justify this limitation. First, these members tend to have the community expertise and community connections needed to represent this constituency's interest on a commission. Second, allowing those without this experience threatens to magnify power imbalances within Black communities, as those with more privileged identities obscure the viewpoints of the most marginalized. Third, because of the racial and economic segregation that persists in American society,[193] this prioritization of members from low-income communities will also aid jurisdictions in determining the relevant community for the purposes of the commission.[194]

iii.   Balancing Power Differentials If Technocrats Play a Role in the Commission

If the power-shifting model adopted includes permanent positions for technocrats, such as policymakers and developers, it would be important to consider and redress the issue of power differentials. Power differentials between technocrat commissioners and community commissioners prompt the risk of co-optation and the fear that marginalized community commissioners might be particularly susceptible to being influenced and strong-armed by non-community commissioners. Promoting participation

---

[193] Residential segregation remains high. *See* LAURA I APPLEMAN, DEFENDING THE JURY: CRIME, COMMUNITY, AND THE CONSTITUTION 78 (2015) (arguing that "[i]ndividual communities in the United States have tended to remain . . . segregated according to race, ethnicity, and class—therefore rendering localized communities more homogeneous"). Also, it is possible to locate the most impacted communities by focusing on most incarcerated zip codes. *See, e.g.*, Caleb Gayle, *Inside the "Most Incarcerated" Zip Code in the Country*, NEW REPUBLIC (Oct. 15, 2019), https://newrepublic.com/articl e/155241/inside-most-incarcerated-zip-code-country.

[194] For non-Black or non-racially marginalized groups, the determination of "the community" will be more difficult. A future work in progress will engage with this problem.

by non-commissioner community members (from the most impacted communities) is one mode of offsetting this risk. Prior to determining the formula of the algorithm, the commission could seek input through stakeholder meetings, public hearings, or community canvassing. These processes would serve to ensure that the commission operates in the interest of the communities most impacted by incarceration and could also have the added benefit of ensuring legitimacy and accountability to the community.

IV. THE POWER-SHIFTING MODEL'S POTENTIAL

If adopted in the pretrial algorithmic governance context, this model could afford a potential resolution to the input problem since it would subject first-order questions around pretrial algorithmic governance to communal consideration and authorization. At the same time, it could engender other benefits that are explored below.

A. *Improvement of Algorithmic Inputs*

Assuming pretrial algorithmic governance is pursued, this model could improve the algorithms that are designed for the pretrial system since it would enable algorithmic systems to benefit from the knowledge of the most impacted communities. This potentiality engages with the Critical Race Theory tradition,[195] which is a scholarly and activist movement that seeks to expose,[196] challenge, and change the way in which race is constructed,[197] deployed, and operationalized to maintain existing power structures. Critical Race Theory supplies the theoretical framework to understand and remedy how race and its intersection with other subordinated statuses[198] operate to privilege white identities over negatively racially marginalized people. I. Bennett Capers has argued that Critical Race Theory provides the vantage point for reckoning with how law constructs race and racial meaning.[199] One of its tenets is that sites of knowledge production and power have

---

[195] It is difficult to define Critical Race Theory succinctly because of the various subgroups of Critical Race Theory that exist. However, it is unified by a set of common questions regarding the power of race and racism within and outside of legal structures. *See* Jasmine B. Gonzales Rose, *Toward a Critical Race Theory of Evidence*, 101 MINN. L. REV. 2243, 2248–49 (2017) (citations omitted).

[196] Though invented, race is ascribed onto individual bodies by a process of racialization, in which certain physical features are imbued with social significance. Kendall Thomas, *The Eclipse of Reason: A Rhetorical Reading of* Bowers v. Hardwick, 79 VA. L. REV. 1805, 1806–07 (1993) ("I have suggested in some of my work in critical race theory that 'race' is a verb, that we are 'raced' through a constellation of practices that construct and control racial subjectivities.").

[197] Angela Onwuachi-Willig, *Policing the Boundaries of Whiteness: The Tragedy of Being "Out of Place" from Emmett Till to Trayvon Martin*, 102 IOWA L. REV. 1113, 1183 (2017) ("[R]ace is defined just as much by stereotypes and the way one behaves in any particular moment and context as it is by the way one looks, and by racially-associated ways of being such as how one dresses, how one styles her hair, how one speaks, and how one votes.").

[198] Kimberlé Crenshaw, *Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color*, 43 STAN. L. REV. 1241, 1244 (1991).

[199] I. Bennett Capers, *Afrofuturism, Critical Race Theory, and Policing in the Year 2044*, 94 N.Y.U. L. REV. 1, 24 (2019).

traditionally privileged white-identified groups, persons, and values to the detriment of those ascribed non-white identities.[200] To counter this problem, Critical Race Theory promotes critical knowledge, embraces racially marginalized people as knowledge producers, and emphasizes the importance of their experiential knowledge in naming and remedying racial injuries.[201] Engaging with this concept, the power-shifting model promotes the utilization of the experiential knowledge of those hailing from the communities that have been most ravaged by the carceral state in the construction, implementation, and oversight of pretrial algorithms, since these groups have knowledge that could counteract and transform the racial hegemonies inherent in the design of current algorithms. Because of their experience with the criminal legal system, members from low-income racially marginalized communities could utilize their unique expertise to develop algorithms designed to mitigate the negative externalities associated with the imposition of incarceration on low-income communities of color. Working alone or alongside technocrats, racially marginalized groups could name, disrupt, and dismantle assumptions that propagate these systems' reproduction of the racial inequities present in the pretrial process.

The actualization of this potentiality could have a range of benefits. Assuming there is a risk component to the algorithms designed, these communities' knowledge could improve this component's predictive accuracy, since these communities might have unique insights into the factors that increase the likelihood of nonappearance and pretrial crime in their neighborhoods—factors that may or may not be the same as factors that are currently utilized. Moreover, communal knowledge could transform risk assessment algorithms to algorithms that provide a more holistic weighing of the risks and harms associated with the defendant's detention before trial. Having experienced firsthand the negative externalities that over-incarceration has had on their communities, community representatives may reject risk-only algorithms in favor of ones that also consider or only consider harms that should count against a defendant's incarceration before trial. An example of such harm-based factors could be the fact that a defendant is the primary caregiver for a minor child. The dual consideration of harm and risk could be achieved by having the algorithm perform two

---

[200] Francisco Valdes, *Outsider Jurisprudence, Critical Pedagogy and Social Justice Activism: Marking the Stirrings of Critical Legal Education*, 10 ASIAN L.J. 65, 70 (2003).

[201] Maria C. Malagon, Lindsay Perez Huber & Veronica N. Velez, *Our Experiences, Our Methods: Using Grounded Theory to Inform a Critical Race Theory Methodology*, 8 SEATTLE J. SOC. JUST. 253, 257 (2009); EDWARD W. SAID, COVERING ISLAM: HOW THE MEDIA AND THE EXPERTS DETERMINE HOW WE SEE THE REST OF THE WORLD 157 (1997) (describing the notion of antithetical knowledge as a "kind of knowledge produced by people who quite consciously consider themselves to be writing in opposition to the prevailing orthodoxy"); RICHARD DELGADO & JEAN STEFANCIC, CRITICAL RACE THEORY: AN INTRODUCTION 9 (2001) ("Coexisting in somewhat uneasy tension with anti-essentialism, the voice-of-color thesis holds that because of their different histories and experiences with oppression, [B]lack, Indian, Asian, and Latino/a writers and thinkers may be able to communicate to their white counterparts matters that the whites are unlikely to know. Minority status, in other words, brings with it a presumed competence to speak about race and racism.").

different assessments, one being a risk assessment and the other being a harm assessment. Because these communities are adversely affected by the over-incarceration of their community members as well as by the release of "high risk" defendants before trial, their control over pretrial algorithmic governance could facilitate the creation of algorithms that account for the full extent of the costs and benefits associated with detaining a defendant before trial.

B. *Democratization of Pretrial Governance for the Most Impacted Communities*

The power-shifting model offers an opportunity to democratize pretrial governance for the most impacted communities. By endowing members from the most impacted communities with the control to resist pretrial algorithmic governance, this model enables communities to voice their opposition to the algorithm project and to an iteration of the pretrial system that is centered on the incarceration of defendants without consideration of the harms that incarceration poses to their communities. Their rejection of this form of governance, if so done, could be read as their resistance to the current status quo and as a desire to prevent the lock-in of the current pretrial system.

At the same time, the power-shifting model could lead to the adoption of an iteration of pretrial algorithmic governance that centers on the most impacted communities, which presents additional democratizing potential. Though traditionally conceptualized as an anti-democratic and technocratic space, this infrastructure controlled by these communities could be transformed into a site that is democratic or even pluralistic, redressing the exclusion and political ostracization that these communities have experienced in pretrial governance.[202] Under this scenario, such an iteration of pretrial algorithmic governance offers a promising platform for low-income racially marginalized people to democratically participate and influence the operation of a pretrial system that has historically muted their voices and viewpoints. Because of how fundamental these algorithms are becoming in the pretrial process and in bail policy,[203] this model would afford members from these communities a path to render bail judges and other bail officials accountable to them and their notion of public safety. Moreover, the model presents a "workaround" to the existing shortcomings of electoral politics that have tended to amplify powerful voices. This would force the majority to hear and contend with the conditions that have promoted the crisis of racialized pretrial incarceration as understood by those inflicted by it. This power-shifting may or may not change outcomes, but it still matters because it has the potential to promote the accountability

---

[202] Rahman, *supra* note 140, at 335; Eaglin, *Constructing*, *supra* note 3, at 66; Collins, *supra* note 3, at 66, 107–10.

[203] Roberts, *Digitizing*, *supra* note 16, at 1716.

and legitimacy of the pretrial system to communities that have traditionally been estranged from it. Beyond pretrial incarceration, if replicated in other parts of the criminal legal system, such a shift of power over algorithms may secure a level of democratic engagement, on the part of the most vulnerable members of society, on a scale that has been unachievable to date.

V. SOME DOUBTS

The power-shifting model could provide a complete redress of the input problem, yet, resultingly, would give rise to objections on the part of algorithm reformers and community advocates. Using pretrial algorithms controlled by the most impacted communities is not a politically palatable or easily implementable solution, since its operationalization is subject to buy-in by communities disproportionately impacted by incarceration, algorithm reformers, bail judges, and the public. The following section elaborates on the difficulties associated with achieving this buy-in, which itself is emblematic of the tension between resolving the input problem on the one hand and the aims of algorithmic governance on the other.

A. *Algorithm Reformers Side*

1. *Dangers of Democratizing Criminal Law Objection*

A major objection to the power-shifting model on the part of algorithm reformers would be that the model places unwarranted faith in the power of communities most impacted by incarceration to transform pretrial algorithms and, resultingly, move the pretrial system away from its carceral and racially stratifying tendencies. A recent articulation of this objection has been authored by John Rappaport. In his article, *Some Doubts About "Democratizing" Criminal Justice*,[204] he doubts that decarceration goals can be achieved through increased lay participation in the criminal legal process, and he instead warns that lay people hold views that are too punitive, divisive, and ill-informed to promote a fairer and more lenient system.[205] Democratization of criminal law, he contends, particularly if designed only to empower racially marginalized people, risks further flaring racial tensions and enlarging the carceral state.[206] Rather than take this chance, "democratizers" should advocate for evidence-based approaches "consistent with democratic values," which would yield a fairer criminal legal system.[207]

This Article's articulation of the power-shifting model is not the form of democratization at which Rappaport's critique is primarily directed since this model aims only to democratize pretrial algorithmic governance and not

---

[204] John Rappaport, *Some Doubts About "Democratizing" Criminal Justice*, 87 U. CHI. L. REV. 711, 759 (2020).

[205] *Id.* at 720.

[206] *Id.* at 808–09.

[207] *Id.* at 810.

the entire pretrial system.[208] Nonetheless, his critique remains applicable because the power-shifting model necessitates increased democratic participation from a segment of the population—the most impacted communities—in the pretrial system. His critique is largely not fatal to the actualization of the model in theory. It is true that members of the Black communities that have been most impacted by incarceration hold heterogeneous and sometimes irreconcilable views about the criminal legal process, which will lead to potentially contentious debates around if and how an algorithm should operate in the pretrial context.[209] However, this is not unique to the most impacted communities. Academics, developers, and technocrats also hold divergent views. The very fact that the pretrial algorithms that are currently in existence are exclusively risk assessment algorithms represent a particular view about the purpose of the pretrial system—a view that is not shared by all. For example, Crystal Yang has suggested the development of a pretrial algorithm designed to perform an analysis that would inform bail judges about the advantages and disadvantages associated with incarcerating a defendant before trial.[210] Her proposal neither promotes community control nor community input, but it demonstrates that the debates about the kinds of algorithms that should be employed in the pretrial system are not settled and will not necessarily be settled if pretrial algorithmic governance remains in the hands of technocrats. Moreover, there is no guarantee that power-shifting would lead to the discontinuance of evidence-based practices or traditional experts; rather, with the power-shifting model, the use of either would at most be subject to community discretion and at a minimum be combined with communal expertise.

Yet, Rappaport's account remains important because it signals some of the political backlash that may arise from shifting power to the most impacted communities. This backlash will be particularly strong among algorithm reformers since the democratization proposed in this Article is in tension with their aims. For many algorithm reformers, part of the appeal of this form of governance stems from its potential to offset the penal populism that is partly responsible for overincarcerating practices, particularly among elected judges.[211] Under this framing, algorithmic governance is supposed

---

[208] Rappaport's critique is aimed at those promoting democratic participation by local communities as the key to fixing the criminal legal system. *See, e.g.*, Laura I Appleman, *Local Democracy, Community Adjudication, and Criminal Justice*, 111 Nw. U. L. Rev. 1413 (2017) (advocating for a more local and democratic conception of criminal justice).

[209] Although there is not unanimity among Black adults regarding the state of the criminal legal system, the majority hold the view that the criminal legal system is less fair to Blacks than to whites. *See* Juliana Menasce Horowitz, Anna Brown & Kiana Cox, Pew Rsch. Ctr., Race in America 2019 11–12 (2019), https://www.pewsocialtrends.org/wp-content/uploads/sites/3/2019/04/PewResearchCent er_RaceStudy_FINAL-1.pdf.

[210] Yang, *supra* note 2, at 1486–88.

[211] *See, e.g.*, Barkow, *supra* note 144 (discussing the role of penal populism in the advent of mass incarceration); Jonathan Simon, *Millennials and the New Penology: Will Generational Change in the*

to offer a powerful counter-majoritarian influence on bail judges' decision-making around the use of pretrial incarceration.[212] Opening up control over pretrial algorithmic governance to the most impacted communities threatens this aim by potentially transforming pretrial algorithmic systems into ones that are more punitive, and, if relied upon by bail judges, might increase the use of pretrial incarceration. Power-shifting does engender this risk. And it does not suffice to counter this concern with the point that there is no evidence that algorithms designed under the power-shifting model will be more punitive. The perception alone threatens the actualization of the entire project.

At the same time, the power-shifting model provides no guarantee that the pretrial algorithms derived under this model will be empirically derived or a product of technocratic expertise, raising fears that the pretrial algorithms produced under this model will be viewed as illegitimate by algorithm reformers and potentially by society as a whole. On the flipside, the fact that this model democratizes pretrial algorithmic governance exclusively for the most impacted communities could jeopardize its political palatability and democratic legitimacy in the eyes of wealthier and whiter communities, who might protest being subjected to an algorithm produced by a commission on which they have no representation. These concerns are not hypothetical. Recent efforts to shift control over policing to local communities, particularly low-income Black and Brown communities, have been met with white backlash.[213] If the policing context is any precursor, overcoming these political and social obstacles may not be possible.

2. *Inaccurate Algorithms Objection[214]*

Another objection that will be raised is that the algorithms designed under the power-shifting model will be "inaccurate." Because the algorithms developed under this model would be derived from communal expertise, the fear will be that the assessments produced by these algorithms would not validly predict a defendant's pretrial risk. A version of this concern is discussed by Sandra Mayson, who warns about attempts to redress algorithmic bias by eliminating racially disparate inputs or altering the weights assigned to such

---

U.S. Facilitate the Triumph of Risk Rationality in Criminal Justice, *in* CRIMINAL JUSTICE, RISK AND THE REVOLT AGAINST UNCERTAINTY 319 (John Pratt & Jordan Anderson eds., 2020) (contextualizing the national consensus around penalism in the 1970s and 1980s).

[212] Wiseman, *supra* note 55, at 417.

[213] Elie Mystal, *The Inevitable Whitelash Against Racial Justice Has Started*, NATION (Aug. 31, 2020), https://www.thenation.com/article/activism/blake-white-people-backlash/; Reggie Jackson, *The Inevitability of a White Backlash to the George Floyd Protests*, MILWAUKEE INDEP. (June 12, 2020), http://www.milwaukeeindependent.com/featured/inevitability-white-backlash-george-floyd-protests/.

[214] This Article uses the term "accuracy" in reference to whether the tool reliably predicts the likelihood of flight or pretrial crime for racially marginalized defendants as compared to white defendants. This issue is generally referred to in the computer literature as "validity" of the tool, but this Article uses "accuracy" since such aligns with the common use of the word.

inputs, without attending to empirical data about nonappearance or arrest for pretrial crime, risk producing inaccurate algorithms.[215]

This objection will be hard to overcome due to the importance that algorithm reformers place on accurate risk prediction. The idea that these algorithmic systems' risk predictions outperform their human counterparts has been a major selling point used by algorithm reformers.[216] One response to this objection could be that currently employed algorithms are not "accurate" since they produce invalid predictions in relation to racially marginalized defendants. But, on a deeper level, this response fails to grapple with the fact that the "inaccurate" objection is tied to the importance that empirics plays in the algorithm project. The truth is that algorithms solely derived from communal expertise or from a blending of communal and technocratic expertise might not yield predictions that are as accurate of nonappearance or pretrial crime as algorithms derived from an empirical model. Moreover, such algorithms may not yield results that could be defined as risk predictions. This only poses a problem if one presupposes that the notion of accurate algorithms will remain tethered to a notion of public safety that is exclusively concerned with the dangerousness that a defendant's release poses to society. A power-shifting model unlocks the potential that algorithms designed within it might pursue a notion of public safety that concerns the safety of the defendant, their family, and their community, alongside society as a whole. Under this model of public safety, accuracy looks different. That difference would be destabilizing both for the algorithm project as well as the pretrial system itself, rendering it hard to imagine the implementation of the power-shifting model.

3. *Rebellion of Bail Judges Objection*

The final objection on the algorithm reformer's side will concern bail judges: Implementing the power-shifting model might cause bail judges to disregard the assessments produced by these algorithms. This concern is not theoretical. There has already been a "rebellion" on the part of bail judges when it comes to heeding the risk predictions of currently employed pretrial algorithms. As Megan T. Stevenson and Jennifer L. Doleac have noted, the influence that a defendant's risk score has on judicial discretion decreases the longer the risk assessment tool is used in the jurisdiction.[217] Moreover, these deviations tend to correlate around race and adversely affect racially marginalized defendants. In a recent study on Kentucky bail decisions, Alex

---

[215] *See* Mayson, *Bias*, *supra* note 3, at 2264–65 (noting that attempts to fix algorithmic discrimination risk inaccurate algorithms).

[216] *See* Sharad Goel, Ravi Shroff, Jennifer Skeem & Christopher Slobogin, *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment*, *in* RESEARCH HANDBOOK ON BIG DATA LAW 10 (Roland Vogl ed., 2021) ("Algorithms typically outperform human judgment in predicting many outcomes, including recidivism.").

[217] Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans* 31, 53 (IZA Inst. of Lab. Econ., Working Paper No. 12853, 2019).

Albright found that bail judges were more likely to override the recommended default for bail bonds (based on risk scores) and impose harsher bond conditions on Black defendants in comparison to similarly situated white defendants.[218] Moreover, Jessica Eaglin's work has shown that sentencing judges are increasingly developing procedural rules to restrict the use of algorithms, which may be foreshadowing for the bail context.[219] Given this climate, algorithms designed under the power-shifting model may exacerbate this trend, particularly in the context of elected judges, who may be primed to view their own assessments as more democratically legitimate than those produced by these algorithms. Of course, this risk may not materialize. Arguably, the fact that these algorithms would be designed by community members may encourage judicial compliance. Moreover, the rebellion of bail judges will not necessarily be abated by the continuation of the current iteration of pretrial algorithmic governance. However, the potential refusal of bail judges to rationalize their decision-making along the lines of the assessments produced by algorithms designed under the power-shifting model poses an important obstacle.

B. *Community Side*

On the other side, the power-shifting model may receive objections from the most impacted communities, particularly from those who reject the algorithm project. For those critics, increased reliance on algorithmic decision-making, regardless of who controls the design, will inevitably reproduce racial stratification. Sean Hill takes this position, contending that risk assessment is inextricably tied to the production of racial tropes that justify the over-detention of Black defendants.[220] In support of his position, he points to algorithms that are currently in existence.[221] His critique is emblematic of the "no-algorithm" position held by some members of the most impacted communities, who resultingly prefer the abolition of the entire pretrial system. Given how algorithmic decision-making has been employed to date, this critique is not easy to dismiss.

Moreover, having an algorithm-based approach to pretrial detention runs counter to the individualized assessment that has been championed by some community organizations. Any algorithm-based approach implicates the age-old debate about rules versus standards.[222] In short, rules require a

---

[218] *See* Alex Albright, *If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions*, 85 Olin Ctr. L., Econ., & Bus. Fellows' Discussion Paper Series 1 , 4 (2019) (There is "suggestive evidence that judges interpret risk score levels differently based on defendant's race."), https://thelittledataset.com/about_files/albright_judge_score.pdf.

[219] Jessica M. Eaglin, *Population Based Sentencing*, 106 Cornell L. Rev. 353 (2021).

[220] Hill, *supra* note 34.

[221] *Id.*

[222] This Article only engages in this important debate to the extent that algorithmic systems apply a rule-based approach to decision-making. For more information about the rules versus standards debate, see generally Frederick Schauer, Playing by the Rules: A Philosophical Examination of

decision-maker to apply a specified outcome to a set of facts.[223] Standards require that a decision-maker perform an individualized assessment as to how a policy should apply to a specific event.[224] The preferability of either approach is context-specific and subject to substantial scholarly debate.[225] The fact that algorithms employ a rule-based approach may garner opposition from those who strongly believe that a defendant's release or detention before trial should be subject to an individualized assessment. It is true that one cost of adopting the power-shifting model is that a defendant may not receive a purely individualized assessment at the bail stage since their bail determination may, in part, be informed by the outcome of an algorithmic system. As the actualization of the power-shifting model is subject to buy-in by the most impacted communities, the no-algorithm position presents an important barrier.

CONCLUSION

The growth of pretrial algorithmic governance presents a troubling challenge for racial justice. Present-day algorithms use factors that entrench racial stratification by promoting harmful stereotypes of Black criminality and Black over-incarceration. By so doing, they operate to maintain the structural disadvantage that denies full citizenship to members of low-income Black communities in this country. As the debate around the racial effects of pretrial algorithmic governance continues, more attention must be paid to the input problem that this form of governance produces and entrenches. By shedding light on this problem, the hope is that this Article adds to conversations underway about the racial justice implications of the algorithm project. Resolving the input problem is potentially realizable and holds latent possibilities for democratization and perhaps the radical reorientation of the pretrial system for the most impacted communities. Yet, the realization of this model would be in direct conflict with the aims and goals of algorithm reformers, rendering it politically impossible. Not only should this result curb enthusiasm for the algorithm project, but it should also cast doubt on the potential of resolving the racial effects of currently

RULE-BASED DECISION-MAKING IN LAW AND IN LIFE (Tony Honoré & Joseph Raz eds., 1991); Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685 (1976).

[223] Kathleen M. Sullivan, *Foreword: The Justices of Rules and Standards*, 106 HARV. L. REV. 22, 58 (1992) (explaining that "[a] legal directive is 'rule'-like when it binds a decisionmaker to respond in a determinate way to the presence of delimited triggering facts. Rules aim to confine the decisionmaker to facts, leaving irreducibly arbitrary and subjective value choices to be worked out elsewhere").

[224] *Id.* at 58–59 (explaining that "[a] legal directive is 'standard'-like when it tends to collapse decisionmaking back into the direct application of the background principle or policy to a fact situation. . . . Standards allow the decisionmaker to take into account all relevant factors or the totality of the circumstances").

[225] Pierre Schlag, *Rules and Standards*, 33 UCLA L. REV. 379, 428–29 (1985) (explaining the epistemological problems with the rules versus standards); Cass R. Sunstein, *Problems with Rules*, 83 CALIF. L. REV. 953, 955, 957 (1995); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 559–68 (1992) (providing an economic analysis of rules and standards).

employed algorithms by measures that do not position the most impacted communities as the designers, implementers, stewards, and controllers of the new technologies of today and tomorrow.