

# The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness

Ben Green  
bgreen@g.harvard.edu  
Harvard University

## ABSTRACT

Risk assessments have proliferated in the United States criminal justice system. The theory of change motivating their adoption involves two key assumptions: first, that risk assessments will reduce human biases by making objective decisions, and second, that risk assessments will promote criminal justice reform. In this paper I interrogate both of these assumptions, concluding that risk assessments are an ill-advised tool for challenging the centrality and legitimacy of incarceration within the criminal justice system. First, risk assessments fail to provide objectivity, as their use creates numerous sites of discretion. Second, risk assessments provide no guarantee of reducing incarceration; instead, they risk legitimizing the criminal justice system's structural racism. I then consider, via an "epistemic reform," the path forward for criminal justice reform. I reinterpret recent results regarding the "impossibility of fairness" as not simply a tension between mathematical metrics but as evidence of a deeper tension between notions of equality. This expanded frame challenges the formalist, colorblind proceduralism at the heart of the criminal justice system and suggests a more structural approach to reform. Together, this analysis highlights how algorithmic fairness narrows the scope of judgments about justice and how "fair" algorithms can reinforce discrimination.

## CCS CONCEPTS

• Social and professional topics → Computing / technology policy; • Applied computing → Law.

## KEYWORDS

risk assessment, criminal justice system, fairness, social justice

### ACM Reference Format:

Ben Green. 2020. The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness. In *Conference on Fairness, Accountability, and Transparency (FAT\* '20)*, January 27–30, 2020, Barcelona, Spain. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3351095.3372869>

## 1 INTRODUCTION

Across the United States, many oft-opposed groups have united around risk assessments as a promising path forward for the criminal justice system: Democrats and Republicans [68], conservative

states [82] and liberal states [120], criminal defense organizations [134] and prosecutors [112]. In turn, risk assessments have proliferated in recent years: in 2017, 25% of people in the U.S. lived in a jurisdiction using a pretrial risk assessment, compared to just 10% four years prior [131]. A 2019 scan of 91 U.S. jurisdictions found that more than two-thirds used a pretrial risk assessment [132].

Risk assessments are mechanisms for identifying potential risks, the likelihood of those risks manifesting, and the consequences of those events [136]. Within the criminal justice system, risk assessments are most widely used in the contexts of pretrial detention (to predict the likelihood that a criminal defendant will fail to appear in court for trial and, in some jurisdictions, will commit a crime before trial) and sentencing and parole (to predict the likelihood that a defendant or inmate will commit a crime in the future). Although actuarial risk assessments have existed within the criminal justice system for several decades, today's tools represent a new generation that incorporates a larger range of risk factors and is often developed through more advanced statistical methods (such as machine learning) [13, 87].

The recent push toward adopting risk assessments is largely motivated by the criminal justice system's current crisis of legitimacy. Scholarship and activism have demonstrated the countless ways in which racism is baked into the criminal justice system's fundamental structure [3, 30, 73, 118, 150]. Through popular books about mass incarceration [3], racial justice movements such as Black Lives Matter, and increased attention to the inequity of policies such as cash bail [31], there is a growing consensus that the criminal justice system is rife with discrimination. Even criminal justice system actors and defenders have acknowledged the need for change. In 2015, more than 130 police chiefs and prosecutors formed a new organization to combat mass incarceration [51]; the following year, the largest police organization in the U.S. apologized for policing's "historical mistreatment of communities of color" [76]. More recently, politicians (including former prosecutors) who formerly embraced "tough on crime" policies have apologized for their actions and championed criminal justice reform [47, 57, 98].

Risk assessments are often hailed as an important tool for addressing some of the criminal justice system's central issues. The theory of change regarding how risk assessments will improve the criminal justice system is grounded in two key assumptions. The first assumption is that risk assessments will mitigate judicial biases by providing "objective" decisions about defendants [36, 68, 120, 121, 164]. With this goal in mind, and following growing evidence that risk assessments and other machine learning models can be biased [4, 10, 123], recent work has focused on developing technical procedures to measure and promote "algorithmic fairness" [11, 23, 29, 91]. Of particular concern is ensuring that risk assessments do not discriminate against blacks relative to whites.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

FAT\* '20, January 27–30, 2020, Barcelona, Spain

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-6936-7/20/02...\$15.00  
<https://doi.org/10.1145/3351095.3372869>

The satisfaction of statistical metrics for fairness has become a central component of evaluating the objectivity of risk assessments.

The second assumption is that risk assessments will promote criminal justice reform. This is expected to occur through objective risk assessments replacing discriminatory policies and reducing incarceration. For example, Senators Kamala Harris and Rand Paul introduced the Pretrial Integrity and Safety Act of 2017, proposing to replace money bail with risk assessments so that pretrial release would be based on risk rather than wealth and so that pretrial release rates would increase [68]. Several states have implemented pretrial risk assessments with these same goals [82, 120]. Many endorsements of evidence-based sentencing are similarly grounded in the goal of reducing incarceration [115, 148].

Supporters of risk assessments draw a clear link between objectivity and criminal justice reform. In its Statement of Principles, Arnold Ventures (the organization behind the Public Safety Assessment (PSA), a pretrial risk assessment used in nineteen states [96]) writes that the goal of its criminal justice reform efforts is to promote “a criminal justice system that dramatically reduces the use of pretrial detention.” Developing the PSA was one of its “earliest investments in pretrial reform,” under the belief that “[p]roviding judges with an objective means to consider only relevant data may counterbalance some [human] biases and lead to fairer pretrial outcomes” [164]. Similarly, the Attorney General of New Jersey described the state’s adoption of “an objective pretrial risk-assessment” as “[o]ne of the most critical innovations undergirding the entire [statewide bail] reform initiative” [127].

Given the centrality of this theory of change to the use of risk assessments, evaluating risk assessments as an approach to criminal justice reform requires interrogating both underlying assumptions.

This analysis requires, as a preliminary step, articulating principles with which to evaluate reform. This is particularly important given that the notion of “criminal justice reform” is itself contested. Criminal justice reform refers broadly to the goal of eliminating or altering policies that lead to mass incarceration and racial injustice. However, there are divergent views about both the causes of and solutions for these challenges. For example, police reform efforts range from focusing on deficiencies in African American male culture (reforms require improving this culture) to the enduring presence of white supremacy and antiblack racism (reforms require structural transformations to U.S. society) [17].

While it is expected that any reform effort will involve multiple visions, the rhetorical flexibility of “criminal justice reform” leads to a significant gap between the expansive change that “reform” suggests and the more minimal shifts that many reformers actually intend. As a result, criminal justice reform rhetoric is often both “superficial”—“most proposed ‘reforms’ would still leave the United States as the greatest incarcerator in the world”—and “deceptive”—many so-called reformers “obfuscate the difference between changes that will transform the system and tweaks that will curb only its most grotesque flourishes” [86].

This paper evaluates reforms based on the extent to which they address the well-documented structural causes of carceral injustice. This is the emphasis articulated by the prison abolition movement, which draws on the slavery abolition movement [38, 110]. Formerly consigned to the outskirts of political discussion, abolition has been

the subject of renewed attention among politicians, the legal academy, social movements, and the media [5, 94, 138]. Prison abolition promotes decarceration with the aim to ultimately create a world without prisons. Recognizing the violence inherent to confining people in cages and to controlling people’s lives through force, abolitionists object to reforms that “render criminal law administration more humane, but fail to substitute alternative institutions or approaches to realize social order maintenance goals” [109]. Nor, however, do abolitionists intend to immediately close all prisons. Instead, abolition is a long-term project to develop “a constellation of alternative strategies and institutions, with the ultimate aim of removing the prison from the social and ideological landscapes of our society” [38]. This involves advocating to end practices such as capital punishment, the use of criminal records in determining access to housing and voting, and the militarization of police [138] and to create alternative practices such as transformative justice, democratic and holistic responses to violence, and increasing resources for education and healthcare [111].

With the aim of structural decarceration in mind, this paper interrogates the theory of change motivating risk assessments. First, building on sociotechnical approaches to objectivity, I demonstrate how the objectivity promised by risk assessments is a chimera: rather than *removing* discretion to create neutral and objective decisions, risk assessments *shift* discretion toward other people and decision points. Second, drawing on legal critiques of rights as tools for achieving just outcomes, I describe how risk assessments are an ill-advised tool for reducing the centrality and legitimacy of incarceration: risk assessments are indeterminate tools that provide no guarantee of reducing incarceration, are made ineffectual by their individualistic conceptions of risk and bias, and are likely to legitimize the structure and logic of criminal punishment. Rather than presenting a viable approach to decarceral criminal justice reform, risk assessments present a superficial solution that reinforces and perpetuates the exact carceral practices that require dismantling.

Risk assessments can, however, be reinterpreted to point toward more substantive criminal justice reform. A proper challenge to risk assessments requires not technical or procedural reforms, but an “epistemic reform” that provides a new interpretation of both risk assessments and the criminal justice system. Thus, having analyzed the impacts of risk assessments *within* the criminal justice system, I turn to questioning what risk assessments tell us *about* the criminal justice system. Returning to the “fairness” of risk assessments, I reinterpret recent results regarding the “impossibility of fairness” [23, 91] as an “incompatibility of equality.” These impossibility results reflect not simply a tension between mathematical metrics of fairness, but instead indicate the fundamental conflict between approaches to achieving justice: *the impossibility of fairness mathematically proves that it is impossible to achieve substantive equality through mechanisms of formal equality*. This epistemic reform challenges the formalist, colorblind proceduralism at the heart of the criminal justice system and provides an escape from the seemingly impossible bind of fairness, exposing an expanded range of possibilities toward achieving criminal justice reform. Moreover, this analysis highlights the severe limitations of fairness as a method for evaluating the social impacts of algorithms, highlighting in particular how algorithmic fairness narrows the scope of judgments about justice and how “fair” algorithms can reinforce discrimination.

## 2 OBJECTIVITY

Although objectivity is often used as a synonym for “science” and “truth,” objectivity is only partially aligned with these terms [37, 129]. The meaning of objectivity comes most directly from its opposition to subjectivity: the goal behind objectivity “is to aspire to knowledge that bears no trace of the knower” [37]. Yet “[t]his ideal of mechanical objectivity, knowledge based completely on explicit rules, is never fully attainable” [129]. The practices followed to produce objectivity are themselves grounded in social norms about what kinds of knowledge are considered objective. These “methods for maximizing objectivism have no way of detecting values, interests, discursive resources, and ways of organizing the production of knowledge,” meaning that “nothing in science can be protected from cultural influence” [67]. Thus, rather than producing knowledge that is truly free from the trace of any people, objectivity represents “knowledge produced in conformity with the prevailing standards of scientific practice as determined by the current judgements of the scientific community” [2].

Objectivity in the form of quantification plays a particularly important role in political contexts rife with distrust, in which officials facing external scrutiny need to depoliticize their actions by “making decisions without seeming to decide” [129]. Particularly in the United States, which is notable for its wariness of individual decision makers, “Techniques such as cost-benefit analysis and risk assessment make it easier to reassure critics within and outside government that policy decisions are being made in a rational, nonarbitrary manner” [77]. Nonetheless, “Study after study and commentary after commentary [have] called attention to the profoundly normative character of risk assessment, showing that it is a far from objective method: indeed, that it is a highly particular means of framing perceptions, narrowing analysis, erasing uncertainty, and defusing politics” [80].

Criminal justice risk assessments exemplify these attributes of objectivity. As concern about discrimination and mass incarceration intensifies, the criminal justice system faces heightened scrutiny. In order to defuse these challenges and depoliticize their actions, criminal justice actors have turned to risk assessments. Practitioners such as probation officers have reported that risk assessments provide defensible grounds for their decisions, making them less vulnerable to criticism [64].

Rather than produce knowledge that lacks any trace of subjectivity, however, risk assessments produce information (and hence outcomes) that is embedded within political norms and institutional structures. Four aspects of risk assessments deserve particular attention as sites where their supposed objectivity breaks down and a great deal of hidden discretion is incorporated: defining risk, producing input data, setting thresholds, and responding to predictions. These sites of discretion exist in addition to the decisions that are inherent to the development of every machine learning model (such as selecting training data and model features [10, 59]) or are external to the risk assessment decision making process itself (such as choosing what interventions should be pursued in response to risk). After this section describes these forms of discretion, the next section will analyze how such discretion hinders risk assessments as a tool for achieving substantive criminal justice reform.

### 2.1 Defining Risk

Risk assessments aim to predict risk, defined as the likelihood of crime. Pretrial risk assessments estimate the risk that a defendant will be rearrested before trial or will not appear for trial; sentencing and parole risk assessments estimate the risk that a defendant or inmate will recidivate. Such predictions typically consider a period of time ranging from six months to two years [107].

Forecasting crime while ignoring the impacts of incarceration causes risk assessments to overvalue incarceration.<sup>1</sup> Releasing someone by definition increases that person’s likelihood to commit a crime in the near future. If crime risk is the primary criterion, then release will always appear to be adverse.

Yet there are many harms associated with incarceration. Pretrial detention significantly increases a defendant’s likelihood to plead guilty, be convicted, and receive long prison sentences [44, 70, 100]. Time spent in prison is associated with negative outcomes including sexual abuse, disease, and severe declines in mental and physical well-being [85, 170]. After being released, former inmates face significant challenges in finding work (a barrier that is stronger for blacks than whites) [124] and suffer disproportionately from depression, serious disease, and death [170]. The families and communities of incarcerated people also face severe hardships [54, 69, 168]. Moreover, because incarceration increases one’s long-term propensity for crime, pretrial detention does not actually reduce future crime [44]. All told, a cost-benefit analysis found that “detention on the basis of ‘risk’ alone can lead to socially suboptimal outcomes” [172].

The emphasis on crime risk also causes risk assessments to absorb the highly racialized meaning of crime. As numerous scholars and lawyers have shown, the types of behaviors that society views with fear and chooses to punish are based in racial hierarchies, such that blackness itself is criminalized [18, 66, 86, 118, 150] and “risk [is] a proxy for race” [66]. As such, risk assessments subsume the racialized concept of crime into a seemingly objective and empirical category that should guide decision making.

### 2.2 Generating Input Data

Some risk assessments rely on information collected by a criminal justice practitioner (e.g., parole officer or social worker) via an interview with a defendant. For example, the widely-used COMPAS risk and needs assessment incorporates information from interviews that include questions such as “Is there much crime in your neighborhood?” [122]. Another risk assessment evaluates individuals along categories such as “Community Disorganization,” “Anger Management Problems,” and “Poor Compliance” [25].

Such questions and categories resist objective answers, turning these assessments into value-laden affairs in which white, Western, and middle-class standards are imposed on defendants [64, 106]. One’s freedom can hinge on these assessments: in 2016, an inmate in New York was denied parole due to a rehabilitation coordinator answering “yes” to the question “Does this person appear to have notable disciplinary issues?” despite the inmate’s lack of a single disciplinary infraction over the prior decade [169].

Recognizing that their evaluations influence the calculations and recommendations of risk assessments, many criminal justice

<sup>1</sup>In practice, risk assessments are based in data about arrests, which typically represents a racially biased measure of crime [45, 108].



practitioners exercise “considerable discretion” in collecting and interpreting information to produce what they see as the appropriate final score [64]. One study found that practitioners ignored or downplayed criminogenic factors in order to produce low risk designations when evaluating minorities who had committed low-level offenses, but interpreted information so as to produce high risk scores when evaluating sexual or violent offenders [64].

## 2.3 Setting Thresholds

Once someone’s risk has been predicted, risk assessments turn the forecasted probability into categories (e.g., low/medium/high [4]) and number ranges (e.g., 1-5 [99]) to be presented to judges. Notably, no prominent risk assessment directly presents probabilities [24] or follows the “intuitive interpretation” [45] of dividing categories across the spectrum of risk (e.g., “low risk” corresponds to 0-33% risk). A related approach is to define risk categories across population percentile (e.g., COMPAS divides the population into ten equal-sized groups, assigning each a score from 1-10 [121]).

In most cases, therefore, the thresholds that determine labels such as “high risk” and recommendations such as “detain” are based in normative judgments about the tradeoffs between reducing incarceration and reducing crime. Jurisdictions implementing the PSA, for example, determine how to define the ranges of low, moderate, and high risk [151]. Although there may be benefits to adapting risk assessments to the local context, doing so introduces a new form of discretion: there is no objective guide for what certain level of risk warrants release or detention. Across risk assessments, the probabilities corresponding to the highest risk categories vary widely and can refer to rearrest rates as low as 3.8% [6, 107]. In turn, public officials often do not actually know how the categories that risk assessments present translate to probabilities of recidivism or failure to appear [86, 92].

These scores and thresholds can have significant impacts on the outcomes of cases. Many jurisdictions directly tie recommendations to the categories defined in the risk assessment [95, 146]. In Kentucky, for instance, the mandatory use of a pretrial risk assessment led to increases in release for low and medium risk defendants and a decrease in release for high risk defendants [152].

Even if a recommendation threshold is set at the outset of reform to promote high levels of pretrial release, it can later be altered to reduce pretrial release. In New Jersey, several defendants accused of certain gun charges were released before trial and then rearrested; the Attorney General’s office then pressured the courts to alter the risk assessment so that it would recommend detention for every defendant arrested for those same gun charges, regardless of that person’s predicted risk [75, 145]. New Jersey soon expanded its detain recommendations to a larger number of offenses [128]. Similarly, in 2017, the United States Immigration and Customs Enforcement (ICE) altered its pretrial Risk Classification Assessment so that it would recommend “detain” in every case [140].

## 2.4 Responding to Predictions

Regardless of how they present predictions, risk assessments typically play a role of decision making aid rather than final arbiter: they provide information and recommendations to judges but do

not dictate the decisions made. Thus, although a common goal behind risk assessments is to eliminate the subjective biases of judges [28, 113, 130, 134, 148, 154, 155, 164], risk assessment implementations allow judges to decide how to respond to the information and recommendations provided.

Many judges use this discretion to ignore risk assessments or to use them in selective ways. In both Kentucky and Virginia, risk assessments failed to produce significant and lasting reductions in pretrial detention because judges tended to override recommendations suggesting release [152, 153]. Judges in Cook County, Illinois diverged from the pretrial risk assessment 85% of the time, releasing defendants at drastically lower numbers than recommended [105]. A juvenile risk assessment faced similar issues: judges frequently overrode the risk assessment when it recommended release, but rarely when it recommended incarceration, leading to a dramatic and “chronic” increase in detention [149]. Similar patterns have been observed in Santa Cruz and Alameda County, California [167].

When they do use risk assessments, judicial decisions are rife with biases. Two experimental studies found that people are more strongly swayed by a risk assessment’s suggestion to increase estimates of crime risk when evaluating black defendants compared to white defendants [60, 61]. Judges in Broward County, Florida have penalized black defendants more harshly than white defendants for being just above the thresholds for medium and high risk [32]. Judicial decisions made with a risk assessment in Kentucky similarly increased racial disparities in pretrial outcomes [1].

It is clear that the first assumption behind risk assessments—that they replace biased discretion with neutral objectivity—does not hold up to scrutiny. Despite being hailed as “objective,” risk assessments shift discretion to different people and places rather than eliminate discretion altogether. Yet the presence of subjective judgment is not itself dispositive as an argument against risk assessments. For if the objectivity sought in risk assessment discourse is impossible, then any reform will rest, to some degree, on discretion. It is therefore necessary to turn to the second assumption motivating risk assessments and evaluate, with these subjectivities in mind, whether risk assessments can spur criminal justice reform.

## 3 CRIMINAL JUSTICE REFORM

Although advocates tend to assume that risk assessments will promote criminal justice reform [68, 127, 164], altering decision making procedures to promote fairness and objectivity does not necessarily reduce incarceration and racial discrimination. Sentencing reform offers a striking case of how the “well-intentioned pursuit of administrative perfection” characteristic of twentieth century civil rights reforms “ultimately accelerated carceral state development” [118]. In 1984, concerned about the racial disparities produced by the judicial discretion to set criminal sentences, Congress passed the Sentencing Reform Act, creating mandatory sentencing guidelines tied to the characteristics of the offender and the offense [101]. This system was designed to constrain judicial discretion and thereby “provide certainty and fairness in meeting the purposes of sentencing” [160]. The reform failed to have the intended impacts, however. The guidelines “set in motion dramatic changes in day-to-day federal criminal justice operations, largely by shifting a massive amount of discretionary power from judges to prosecutors”

[101]. The result was a “punitive explosion” that increased both incarceration and racial disparities [101].

Similar reforms throughout U.S. history have centered on the expansion of rights as a mechanism to promote fair procedures. These rights-based reforms often did not actually notably improve outcomes: for instance, schools remained segregated and unequal well after the Supreme Court deemed school segregation unconstitutional in *Brown v. Board of Education* [157]). U.S. legal scholars in the twentieth century therefore developed the “critique of rights”—a critique of rights-based reforms and discourse in mainstream legal thought. Advanced by scholars such as Duncan Kennedy [88] and Mark Tushnet [156, 157], the critique of rights revolves around five assertions: 1) Rights are less effective at spurring progressive social change than commonly assumed, 2) The impacts of rights are indeterminate, 3) The discourse of rights abstracts away the power imbalances that create injustice, 4) The individualistic discourse of rights prioritizes individual freedom over social solidarity and community well-being, and 5) Rights can impede democracy by reinforcing undemocratic relationships and institutions [22].

Today’s appeals to risk assessments mirror historical appeals to rights: like rights reforms such as the right to a lawyer, the introduction of risk assessments into bail and sentencing is intended to produce a fair and neutral process for criminal defendants [68, 120, 164]. This suggests that risk assessments should be interrogated against the critique of rights. Doing so, I show that risk assessments suffer from the same core limitations as rights: they are indeterminate, individualistic, and legitimizing.

### 3.1 Indeterminate

Although risk assessments are often hailed as objective, a great deal of subjective judgment resides under the surface of these tools. This discretion can dramatically alter the use and impacts of risk assessments. In this sense, risk assessments are indeterminate: the adoption of risk assessments provides little guarantee that the intended social impacts will be realized.

Indeterminacy is a common feature of decision making processes grounded in rules and procedures [157]. Procedural reforms often fail to generate the intended outcomes because they use technical means to achieve normative ends. Achieving the desired outcomes requires a particular use of the tool or process, yet nothing about the procedures guarantee that such use will arise in practice. As noted in the critique of rights, the adoption of a progressive law provides little guarantee of the political outcomes seemingly connected to that law; instead, broader social circumstances largely dictate how that law will be wielded, interpreted, and applied. And “if circumstances change, the ‘rule’ could be eroded or [even] interpreted to support anti-progressive change” [157].

Risk assessments are unreliable as tools for reducing incarceration because they depend on the social and political circumstances of their use. Risk assessments are embedded in the criminal justice system, in which the structural and political incentives largely favor punitive and carceral policies [3, 16, 18, 153]. Thus, to the extent that the types of subjectivity described in Section 2 manifest in risk assessments, such discretion typically resists decarceral goals. Definitions of risk emphasize incarceration as a way to reduce crime while ignoring the significant harms of incarceration. Interviews

and evaluations allow white and middle-class assumptions (which typically associate blackness with crime, aggression, and a lack of innocence [55, 56, 119, 135]) to influence judgments about defendants and inmates. The practice of defining thresholds allows for people with low probabilities of rearrest to be labeled “high risk” and for recommendations to be altered to reduce how many people are released. Judicial responses to risk assessments exacerbate racial disparities and diminish release rates.

These forms of discretion make the impacts of risk assessments brittle and prone to political capture. Achieving decarceral outcomes through risk assessments requires particular behaviors and circumstances which the criminal justice system is generally not amenable to. As a result of this indeterminacy, risk assessments provide no guarantee of reducing incarceration and in fact are often wielded in ways that resist decarceral outcomes. Yet because of the discourse that positions risk assessments as a tool for reform, even ineffective implementations may enhance perceptions of fairness and reduce the political will for more systemic changes.

### 3.2 Individualistic

Risk assessments are based on individualistic conceptions of both risk and bias that lead to individualistic and ineffectual remedies for racial discrimination and mass incarceration.

Risk assessments treat risk at the level of individuals, defining risk in terms of someone’s likelihood to be arrested in the future. This approach treats risk as a measure of difference across individuals—an objective and static fact of identity—rather than as a social category defined through social norms (what is considered a crime) and relations (why certain people commit and are punished for those crimes).

Although numerous social markers of difference are accepted as “intrinsic” and “natural,” many of these categories emerge from social arrangements that imbue those comparisons with meaning and importance [114]. In particular, “difference” becomes salient when “a more powerful group assigns meaning to a trait in order to express and consolidate power” [114]. For example, “[w]omen are compared with the unstated norm of men, ‘minority’ races with whites, [and] handicapped persons with the able-bodied” [114]. Addressing difference equitably requires not providing special treatment (whether ameliorative or punitive) to “different” individuals, but altering the relationships and institutions that structure these categories [114].

Risk assessments focus on individual-level risk, leading them to suggest individual-level interventions. Calculating each person’s risk differentiates risk factors across members within a population, but obscures the structural factors that shape the distribution of risk itself [133]. In other words, risk assessments make legible the idea of high-risk *individuals* rather than high-risk *populations*. As a result, risk assessments justify individualistic responses: most notably, incarcerating high-risk people. Yet it is precisely population-level reforms such as improving access to housing, healthcare, and employment that are most likely to reduce crime risk and improve well-being across the population [65, 71, 126, 143, 144].

Because risk assessments focus on individuals, they can entrench historical injustice by failing to recognize changing social circumstances. Risk assessments (as with all machine learning) assume

that population characteristics are constant, such that factors producing certain outcomes in the past will produce those outcomes at the same rates in the future. Even if jurisdictions enacted reforms that reduce crime, risk assessments would be blind to these new circumstances. In turn, risk assessments would overestimate crime and recommend incarceration for individuals whose crime risk has decreased. Following interventions such as text messages that remind defendants to appear in court, risk assessments have produced “zombie predictions” that overestimate risk because they fail to account for the risk-reducing benefits of these reforms [92]. And because incarceration increases the likelihood of crime after someone is released [35, 40, 165], these false positive predictions will exacerbate the cycle of recidivism and incarceration that risk assessments are meant to remedy.

Risk assessments suffer from a similarly individualistic approach to bias: they diagnose bias as a behavior exhibited by individuals, typically due to implicit bias. Risk assessments are therefore designed to replace the discretion of biased judges with “objective” algorithmic predictions [28, 113, 130, 134, 148, 154, 155, 164].

Yet this emphasis on the bias of individuals overlooks the policies and institutions that structure racial hierarchies. Discrimination and oppression are produced not simply by people making biased judgments, but through laws and institutions that systematically benefit one group over another [3, 86]. Diagnosing discrimination as the product of discretion and bias “displace[s] questions of justice onto the more manageable, measurable issues of system function” [118], thus “obscur[ing] the larger structural aspects of racism” and “draining attention and resources away from other approaches to framing and addressing racism” [83].

By focusing on judicial decisions as the source of discrimination, risk assessments shroud the social structures and power dynamics behind racial discrimination. They obscure the need to transform policies and institutions in order to achieve racial equity, instead suggesting that discrimination can be remedied by altering decision making procedures. Attempts to address racial oppression that focus solely on the bias of individual decision makers serve to legitimize and reinforce that oppression.

### 3.3 Legitimizing

Despite being implemented under the banner of criminal justice reform, risk assessments naturalize and legitimize carceral logics (e.g., risky defendants should be held before trial) and practices (e.g., determining which defendants are “risky”).

Across domains, reforms that address the salient aspects of an injustice rather than the underlying causes and conditions of that injustice can legitimize those underlying structures. For instance, efforts to eradicate war crimes such as torture without challenging war itself have “tolerated the normalization of perpetual, if more sanitary, war” [117]. Closer to criminal justice reform, diversity and implicit bias trainings present a notable example of how reforms aimed at preventing discrimination can legitimize social arrangements that produce inequality. Numerous studies have found these trainings to be ineffective at improving diversity or reducing bias [83]. Instead, by creating “an illusion of fairness” that “legitimizes” existing social arrangements” [84], the “formal bureaucratic procedures may reproduce inequality rather than eradicating it” [89].

Individualistic and procedural reforms are particularly prone to legitimization. When it comes to legal rights, “progressive victories are likely to be short-term only; in the longer run the individualism of rights-rhetoric will stabilize existing social relations rather than transform them” [157]. This observation, that winning a legal battle can rely on principles (such as individualism) that hinder long-term efforts for structural transformation, is known as “losing by winning” [157]. With regard to criminal rights (such as the guarantee that every criminal defendant be provided with an attorney), “procedural rights may be especially prone to legitimate the status quo, because ‘fair’ process masks unjust substantive outcomes and makes those outcomes seem more legitimate” [20]. The enactment of such rights “makes it more work—and thus more difficult—to make economic and racial critiques of criminal justice” [20].

Risk assessments exemplify an individualistic and procedural reform as well as the limits of this approach. Risk assessments focus on decision making procedures: their primary concern is not that incarcerating people is wrong, but that decisions about which individuals to incarcerate should be reached more empirically and objectively. This represents a narrow vision of reform, one that attempts to measure risk without bias or error while upholding the notion that incarceration is an appropriate response to “high-risk” individuals. In other words, risk assessments focus reform efforts on decisions about individuals while overlooking the structures shaping that decision, who is subject to it, and what its impacts are. Although presented under the banner of reform, this type of “[a]dministrative tinkering does not confront the damning features of the American carceral state, its scale and its racial concentration” [118]. Instead, by tweaking surface-level decisions and providing them with a semblance of neutrality and fairness, risk assessments are likely to sanitize, legitimize, and perpetuate the criminal justice system’s carceral and racist structure. From the perspective of decarceration and racial justice, the enactment of risk assessments represents a clear example of “losing by winning.”

This process of legitimization can be seen most clearly with regard to preventative detention (detaining a criminal defendant before trial due to concerns about crime risk). The practice was not deemed constitutional until the 1987 U.S. Supreme Court case *United States v. Salerno* [9, 92, 172]. Yet today the practice of preventative detention—which Supreme Court Justices Marshall and Brennan deemed “incompatible with the fundamental human rights protected by our Constitution” [162]—is being legitimized as a central aspect of “modern” [120] and “smart” [163] criminal justice reforms based on risk assessments. Through such logic, the use of risk assessments as tools for reform “conced[es] Salerno” and “ratifies recent erosions of the fundamental rights of the accused” [92].

These three attributes of risk assessments—indeterminate, individualistic, and legitimizing—demonstrate the flaws of the assumption that risk assessments will promote criminal justice reform (at least with regard to any notion of reform that involves reducing the centrality of punishment and incarceration). These tools are poorly suited to the task of combatting carceral practices and logics. Despite being presented as a valuable mechanism for racial justice, risk assessments are akin to the many components of criminal justice reform today that are oriented around “the margins of the problem without confronting the structural issues at its heart” [86].



Thus far I have shown that the theory of change behind risk assessments is deficient: neither of the two core assumptions, regarding objectivity and reform, withstand close inspection. The question that remains is what this suggests for criminal justice reform efforts: How can risk assessments be challenged in a manner that facilitates a path toward more systemic reform?

## 4 EPISTEMIC REFORM

The movement for risk assessments derives not simply from the presence of particular technologies (i.e., big data and machine learning), but from a particular understanding of social challenges as technological in nature and amenable to technological solutions. As pressure mounts for criminal justice reform in an era of “technological solutionism” [116], “technochauvinism” [14], and “tech goggles” [59], what has emerged is a “sociotechnical imaginary”—a collective vision of a desirable future attainable through technology [81]—that casts criminal justice adjudication as prediction tasks, ones that algorithms can perform better than humans. Holding together these imaginaries and technologies is “co-production,” which describes how “the ways in which we know and represent the world (both nature and society) are inseparable from the ways in which we choose to live in it” [78]. Through co-production, it is often new technological *discourses* rather than new technological *artifacts* that provide a sense of order in the face of instability [79]. Yet these discourses, however secure and widespread they may appear, are not static: altering forms of knowledge “can function as strategic resources in the ongoing negotiation of social order” [104].

This emphasis on discourses in addition to artifacts can inform the appropriate responses to the false promises of risk assessments. The dangers of risk assessments are not the result of poor implementation, but are instead inherent to the sociotechnical imaginary that treats criminal justice as a set of prediction problems. Under this framing, attempts to generate “better” (i.e., fairer and more accurate) risk assessments are unlikely to reduce these tools’ fundamental harms. Rather than calling for unbiased risk assessments, then, a more fruitful path to diminishing carceral logics and practices is to present an “epistemic challenge” [166] to the sociotechnical imaginary around risk assessments. Such an “epistemic reform” can shift our focus from evaluating risk assessments through the lens of the criminal justice system to evaluating the criminal justice system through the lens of risk assessments. Doing so can point the way toward more effective criminal justice reforms.

### 4.1 Reinterpreting the “Impossibility of Fairness”

Although it is common to discuss risk assessments and judges using the same language of bias—and even to directly compare their biases [11, 155]—“bias” has distinct meanings across these two contexts. The bias of a judge speaks to something individual: the implicit and explicit biases that influence a specific person’s decisions. The “bias” of a risk assessment, on the other hand, speaks to something structural: the ways in which different groups of people are systematically filtered to different outcomes.

To understand this distinction, it is necessary to distinguish between two causes of algorithmic “bias”:

- (1) **Human Bias:** The first form of “bias” occurs when an algorithm is trained on the decisions of biased humans—a type of “garbage in, garbage out.” For instance, a risk assessment would be subject to Human Bias if its training data overestimates the recidivism rates of black defendants due to over-policing in black neighborhoods. Because this algorithm would be learning to reproduce human biases, it seems appropriate to refer to its decisions as “biased” and to make the comparison with human bias.
- (2) **Population Inequity:** The second form of “bias” occurs when an algorithm is trained on population-level disparities—a type of “inequity in, inequity out.” For instance, a risk assessment would be subject to Population Inequity if its training data reflects (beyond any distortion from Human Bias) that black defendants are more likely than white defendants to recidivate. Because this algorithm would be learning to reproduce social outcomes that are the product of historical oppression, its discrimination is not akin to the bias of human decision makers.

Failing to distinguish Human Bias from Population Inequity can hinder efforts to understand and reduce algorithmic discrimination.<sup>2</sup> Population Inequity is most directly related not to the biases of judges or other people, but to “the racial inequality inherent in *all* crime prediction in a racially unequal world” [108].

To see this challenge of making fair predictions in an unequal society, consider the recent statistical results regarding the “impossibility of fairness” [23, 91]. The results concern two metrics for evaluating fairness. The first metric is calibration, which states that predictions of risk should reflect the same underlying level of risk across groups (i.e., 50% risk should mean a 50% chance of rearrest whether the defendant is black or white).<sup>3</sup> Calibration is akin to colorblindness. The second metric is error rate balance, which states that false positive and false negative rates should be equal across groups. Given these two metrics, the impossibility of fairness shows that if two groups have different rates of an outcomes, then it is impossible for predictions about those groups to both be calibrated and have balanced errors. In the context of risk assessments, this means that given higher crime rates among black defendants than white defendants, it is impossible for a risk assessment to make calibrated predictions of risk without having a higher false positive rate (and lower false negative rate) for black defendants.

From this perspective, risk assessments appear to be situated within an “impossible” set of tradeoffs [63]. In turn, the impossibility result is often interpreted as a defense of calibrated decision making. When ProPublica demonstrated COMPAS’ disproportionate false positive rates for black defendants [4], Northpointe (the company, now known as Equivant, that created COMPAS) refuted that higher recidivism rates among blacks explained the disparity and thereby absolved them from accusations of racial bias. They wrote, “This pattern does not show evidence of bias, but rather is a natural consequence of using unbiased scoring rules for groups that happen to have different distributions of scores” [43]. Other scholars have

<sup>2</sup>Another paper has made a similar distinction, between 1) “racial distortions in past-crime data relative to crime rates” and 2) “a difference in crime rates by race” [108]. The two phenomena can also coexist: they are distinct but not mutually exclusive.

<sup>3</sup>A related measure is predictive parity, which states that the outcome rates among people labeled “high risk” should be the same across groups.

similarly pointed to the incompatibility of fairness metrics to dispel claims that algorithms are discriminatory [11, 28, 155].

Yet the problem of discrimination is not so neatly resolved by reference to the underlying base rates: the disparities in these population-level statistics are *themselves* the product of discrimination. African Americans do not just “happen to have different distributions of scores”—blackness itself is criminalized [18, 66, 86, 118, 150] and blacks have been subjected to myriad forms of oppression (including redlining and segregation [141], the war on drugs [3, 86], and severe underfunding of schools [46]) that contribute to increasing crime [93, 97, 139, 142].

Notions of fairness in risk assessments generally fail to consider such context, however. In both research and practice, calibration is the typical instantiation of fairness [29, 43, 108]. Yet calibration strives for accurate predictions of risk, regardless of the factors structuring that risk. Risk assessments thus overlook the social conditions behind racial disparities, striving to accurately identify risk without interrogating whether that notion of risk is appropriate, why some people have high levels of risk, or whether incarceration is an appropriate response for high-risk people. Rather than being blind to color, calibrated risk assessments are blind to structural oppression.

Consider the gold standard: a hypothetical risk assessment that predicts with perfect accuracy whether each person will recidivate.<sup>4</sup> Such a risk assessment would satisfy all three metrics of fairness that are typically in tension [91]. The impossibility would disappear. Yet this risk assessment would *still* disproportionately label blacks as “high risk” compared to whites—not because of Human Bias, but because of Population Inequity: due to discrimination and the racialized meaning of “crime” and “risk” [18, 66, 86, 118], African Americans *are* empirically at higher risk to commit crimes [27, 143, 147, 159]. In other words, because “[r]acism is not a mistake, not a matter of episodic, irrational behavior” [41], eliminating inaccurate predictions will not eliminate racist predictions.

Herein lies the danger of overlooking Population Inequity: accounting only for Human Bias, even with a “perfect” risk assessment, would still subject blacks to higher rates of incarceration than whites. This “fair” algorithm launders the products of historical discrimination into neutral and empirical facts, in turn reinforcing this discrimination by punishing African Americans for having been subjected to such criminogenic circumstances in the first place.

This conflict in algorithmic fairness between Human Bias and Population Inequity speaks to a more fundamental tension between notions of equality: formal equality and substantive equality. This tension runs throughout debates in areas ranging from equality of opportunity [49] to antidiscrimination [33] to big data [10]. Formal equality emphasizes equal treatment or equal process: similar people should be treated similarly. Substantive equality emphasizes equal outcomes: groups should obtain similar outcomes, even if that requires accounting for different social conditions between groups. In the U.S. legal context, disparate treatment is grounded in notions of formal equality (or anticlassification) while disparate impact is grounded in notions of substantive equality (or antisubordination).

By ensuring that individuals who have similar levels of risk are treated similarly, calibration expresses the logic of formal equality.<sup>5</sup> In this sense, calibration aims to account for Human Bias: it strives for predictions that reflect one’s actual level of risk, untainted by distortions. Alternatively, by ensuring that groups are similarly affected by false predictions, error rate balance expresses the logic of substantive equality [108]. In this sense, error rate balance aims to account for Population Inequity: it strives for risk predictions that do not disproportionately harm one group more than another, regardless of the underlying distributions of risk.

With these parallels in mind, the epistemic reform becomes possible: the “impossibility of fairness” can be reinterpreted as an “incompatibility of equality.” Because calibration is a measure of formal equality and error rate balance is a measure of substantive equality, the impossibility result can be restated as a tradeoff between formal and substantive equality: *the impossibility of fairness mathematically proves that, in an unequal society, decisions based in formal equality are guaranteed to produce substantive inequality*. Although the impossibility of fairness is typically taken to indicate that disparate outcomes are the mere byproduct of fair risk assessments [11, 28, 43, 155], this reframing highlights the opposite: disparate outcomes are the inevitable product of colorblind risk assessments in an unequal society.

Notably, it is precisely the desire for objectivity that grounds risk assessments in formal equality and makes them unable to generate substantive equality. Dominant notions of racial equality based in colorblindness developed from a desire for neutrality and objectivity, in direct opposition to more radical calls for racial justice from the black nationalist movement [125]. Because colorblindness entails “the refusal to acknowledge the causes and consequences of enduring racial stratification” [118], it “creates and maintains racial hierarchy much as earlier systems of control did” [3]. Thus, just as the law “will most reinforce existing distributions of power” when it is “most ruthlessly neutral” [102], risk assessments will most entrench racial injustice when they are most (seemingly) objective.

## 4.2 Implications for Criminal Justice Reform

Statistical arguments that articulate these tensions between formal and substantive equality can challenge fundamental inequities in the criminal justice system. The Supreme Court confronted this issue in the 1987 case *McCleskey v. Kemp*, in which Warren McCleskey, an African American convicted of killing a white police officer, was sentenced to the death penalty in Georgia [161]. McCleskey challenged this verdict with statistical evidence of structural inequality: the death penalty was disproportionately applied in murder cases with black defendants and white victims [7].

Despite this evidence, the Supreme Court affirmed the death penalty ruling. It argued that the statistical evidence failed to demonstrate deliberate racial bias in McCleskey’s case. In the majority opinion, Justice Lewis Powell wrote, “a defendant who alleges an equal protection violation has the burden of proving the existence of purposeful discrimination. [...] McCleskey must prove that the decisionmakers in *his* case acted with discriminatory purpose” [161].

<sup>4</sup>For the sake of this example, suppose that the training data and outcomes reflect an accurate and unbiased measure of crime (i.e., there is no Human Bias).

<sup>5</sup>Although Mayson characterizes calibration as a disparate impact metric, I argue that it more closely aligns with the disparate treatment logic of ensuring that people with the same risk receive the same score (an equivalence that Mayson acknowledges) [108].



Powell provides a formal equality analysis: the outcome is legitimate as long as McCleskey was not subject to intentional discrimination.

Powell further justified this outcome by arguing that acknowledging substantive inequality in the face of formal equality would cause the entire structure of criminal law to crumble. Recognizing that “McCleskey challenges decisions at the heart of the [...] criminal justice system,” he wrote,

In its broadest form, McCleskey’s claim of discrimination extends to every actor in the Georgia capital sentencing process, from the prosecutor who sought the death penalty and the jury that imposed the sentence, to the State itself that enacted the capital punishment statute and allows it to remain in effect despite its allegedly discriminatory application. We agree with the Court of Appeals, and every other court that has considered such a challenge, that this claim must fail. [161]

The epistemic reform regarding risk assessments can embolden the discourse that Justice Powell recognized as an existential threat to the criminal justice system. Reinforcing the work of other scholars who have articulated the tensions between formal and substantive equality with regard to race and sex [33, 42, 103, 118, 125], the impossibility of fairness provides a mathematical proof of the inherent conflict between formal equality procedures and substantive equality outcomes in an unequal society. Failing to acknowledge the legacy of historical oppression will allow even “fair” risk assessments to perpetuate racial inequity. As Justice William Brennan remarked in his dissent in *McCleskey*, “we remain imprisoned by the past as long as we deny its influence in the present” [161].

Furthermore, the systematic nature of risk assessments may allow the incompatibility of equality to carry more force than the statistical evidence in *McCleskey*. In *McCleskey*, the Supreme Court argued that some variation in the outcomes of similar cases results from the “discretion [that] is essential to the criminal justice process” [161]. Risk assessments are specifically designed to replace judicial discretion with standardized objectivity, however. Moreover, algorithmic discrimination reflects not the bias of an individual but the systematic filtering of different groups into disparate outcomes. To the extent that judgments are standardized by risk assessments, then, statistical evidence of racial disparities could become increasingly difficult to defend on procedural grounds of discretion and could instead be recognized as reflecting structural discrimination.

Such evidence on its own will not function as a “*deus ex data*” that prompts a restructuring of the criminal justice system. One lesson to be learned from *McCleskey* is that social scientific evidence may do very little to persuade courts to accept claims of discrimination [19]. Indeed, despite ProPublica’s evidence that COMPAS disproportionately labeled black defendants with false positive predictions of recidivism, the Wisconsin Supreme Court upheld the use of COMPAS at sentencing in *State v. Loomis* [171].

Achieving decarceral reform therefore requires emphasizing the *interpretation*—not just the *design*—of risk assessments as a site of contest. A focus on reframing notions of crime and criminal justice has long been at the heart of fights for racial justice. In *Black Feminist Thought*, Patricia Hill Collins writes that “activating epistemologies that criticize prevailing knowledge and that enable us to define

our own realities *on our own terms*” is essential to empowering black women [26]. Prison abolition similarly aims to dismantle carceral discourses and to create alternative, emancipatory ones. Prisons are so ingrained in culture and common sense that “it requires a great feat of the imagination to envision life beyond the prison” [38]. The path toward decarceration therefore requires society “to counter criminological discourses and knowledge production that reify and reproduce carceral logics and practices” [15].

Risk assessments are often hailed in ways that reify and reproduce carceral logics and practices. Yet by expanding the scope of analysis, it is possible to reinterpret risk assessments to demonstrate the limits of dominant anti-discrimination frameworks and to identify a path toward more structural criminal justice reform. The emphasis on substantive equality enables a reform approach that avoids the seemingly intractable bind presented by the impossibility of fairness and the false choice between implementing risk assessments and doing nothing [11, 90, 108]. For when faced with decisions that significantly structure its subjects’ lives, the answer is not to optimize the formal fairness of that decision but “to renovate the structure [of the decision] itself, in ways large and small, to open up a broader range of paths that allow people to pursue the activities and goals that add up to a flourishing life” [49].

There are countless opportunities to renovate the structure of criminal justice decisions and thereby escape the “impossible” choices of risk assessments. Criminal justice institutions can change what interventions are made based on risk assessments, responding to risk with support rather than punishment [8, 62, 108]. Reducing pre-trial detention and mandatory minimums [52] (reforms which polls suggest are popular [12, 50, 53]) can further diminish the harms and scope of risk assessments. The gaze of risk assessments can be turned from defendants to the actors and institutions that comprise the criminal justice system [21, 34], enabling a more structural view of the system’s operations. Governments can implement policies that reduce the risk of general, pretrial, and inmate populations [39, 71, 92, 158], thus diminishing the role for punitive responses to risk. The logic behind such reforms is not to reject risk assessments in favor of the status quo, but to reject the structures underlying risk assessments in favor of decarceral and non-punitive structures.

## 5 DISCUSSION: ALGORITHMIC FAIRNESS AND SOCIAL CHANGE

Despite their widespread support, risk assessments are based in a deficient theory of change: they provide neither objectivity nor meaningful criminal justice reform. Risk assessments bear no guarantee of reducing incarceration—instead, they are more likely to legitimize the criminal justice system’s carceral logics and policies. Yet because support for risk assessments emerges in part from the sociotechnical imaginary that sees all problems as solvable with technology, critiques that articulate the technical limits of risk assessments will likely be met by calls for “better” risk assessments. It is therefore necessary to pursue an “epistemic reform” that challenges the *discourses* rather than the *technical specifications* of risk assessments. The impossibility of fairness can be reinterpreted as an incompatibility of equality, demonstrating how mechanisms of formal equality in an unequal society lead to substantive inequality. Seen in this light, risk assessments demonstrate the limits of

formalist, colorblind proceduralism and suggest a more expansive and structural approach to criminal justice reform.

These arguments highlight the myopia of “fairness” as a framework for evaluating the social impacts of algorithms. Although researchers have tended to equate technical and social notions of fairness, fairness in its myriad and conflicting meanings cannot be reduced to a single mathematical definition that exists in the abstract, apart from social, political, and historical context [63]. Guaranteeing these technical conceptions of fairness is therefore drastically insufficient to guarantee—or even reliably promote—just social outcomes. Two issues in particular stand out.

First, algorithmic fairness sidelines the social contexts in which decision making occurs. It treats fairness as a matter of making accurate predictions but does not interrogate the structures behind why certain people are prone to the outcome being predicted or what actions are taken based on predictions. With some exceptions [10, 11, 29, 48], algorithmic fairness debates and metrics hinge on comparing false predictions across groups [4, 11, 72, 91, 108], the implication being that a perfectly accurate model would eliminate the core problem of unfairness. Indeed, recent scholarship asserts that “[t]he most promising way to enhance algorithmic fairness is to improve the accuracy of the algorithm” [72] and that “[t]he largest potential equity gains may come from simply predicting more accurately than humans can” [90].

Although there are fairness benefits to be achieved through improving the accuracy of predictions, the emphasis on accuracy reveals how algorithmic fairness is primarily concerned with Human Bias rather than Population Inequity. Accurate predictions about an unequal society are typically seen as fair. Yet even a “perfect” risk assessment will reinforce the racial discrimination that has structured all aspects of society. As such, algorithmic fairness narrows the scope of judgments about justice, removing structural considerations from view. In this way, algorithmic fairness “mirror[s] some of antidiscrimination discourse’s most problematic tendencies,” most notably the “fail[ure] to address the very hierarchical logic that produces advantaged and disadvantaged subjects in the first place” [74]. Avoiding the perpetuation of historical harms through algorithms “will often require an explicit commitment to substantive remediation rather than merely procedural remedies” [10].

Second, algorithmic fairness fails to account for the trajectory of social change facilitated by algorithms. Although often intended to improve society, algorithms can—even when satisfying fairness criteria—perpetuate or exacerbate inequities. Evaluations of fairness do not consider the harms of an individualistic approach to reform, the potential of algorithmic decision making to legitimize unjust systems, or the dangers of conceiving decision making and reform as technical projects. Instead, an algorithm’s fairness is treated as determinative of it having fair social impacts; as long as risk assessments can lead to more accurate or fair decisions, the thinking goes, they are a step in the right direction [11, 58, 155].

Yet creating a more equitable society is not simply a matter of having algorithms generate marginally improved outcomes compared to the status quo—it requires responding to social challenges with holistic responses that promote egalitarian structures and outcomes in both the short and long term [58, 59]. As “an aspirational ethic and a framework of gradual decarceration,” abolition aims not to make the criminal justice system more humane while retaining

its essential structure, but to reduce the need for (and ultimately eliminate) carceral responses to social disorder [110].

Responsibly developing and evaluating algorithms as tools for social progress requires new methods based in the relationship between technological interventions and social outcomes. First, recognizing the indeterminacy of procedural reforms, reform advocates should avoid deterministic assumptions about the impacts of technology. Rather than viewing technology as a discrete agent of predictable change, reformers should consider the potential for unexpected impacts and should ground any algorithms used within circumstances conducive to reform. For instance, drawing on approaches to limiting legal indeterminacy, the implementation of algorithms could be tied to “sunset provisions” that condition ongoing use of the algorithm to approval based on the results of algorithmic impact assessments [137]. Second, to counter the harms of individualistic decisions and logics, computer scientists must develop new methods that recognize and account for the structural conditions of discrimination, oppression, and inequality. Third, rather than developing tools that are likely to streamline and legitimize existing systems, algorithm developers should thoughtfully consider what interventions will actually be effective at promoting the desired social outcomes. In many cases, typical algorithmic “solutions” may be counterproductive compared to alternative algorithmic approaches or non-algorithmic reforms.

The challenges *raised* by questions of algorithmic fairness are not—and must not be—limited to the scope of analysis *presented* by algorithmic fairness. Algorithmic decision making raises fundamental questions about the structure of institutions and the types of reform that are appropriate in response to injustice. Yet as currently constituted, algorithmic fairness narrows these debates to the precise functioning at the decision point itself. This approach overlooks and legitimizes the context that gives structure and meaning to the decision point. In turn, it leads down a path toward dilemmas that, within this scope, appear intractable. Escaping these false choices requires that “we question [our] assumptions and try to look at the issues from another point of view” [114]. Approaching algorithms as sociotechnical imaginaries rather than as discrete technologies enables this expanded scope of analysis. By highlighting the entire context surrounding algorithms as subject to reimagination and reform, this approach avoids the trap of false dilemmas and makes possible more substantive change. Engaging in this manner with today’s complex socio-legal-technical environments will inform new paths for algorithms and for reform, in the criminal justice system and beyond.

## ACKNOWLEDGMENTS

Thank you to the FAT\* reviewers and track chair and to Salomé Viljoen for their thorough and thoughtful comments on this paper. I am grateful to Sheila Jasanoff and Jeannie Suk Gersen for introducing me to many of the ideas behind this paper. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1745303. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

- [1] Alex Albright. 2019. If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions. *The John M. Olin Center for Law, Economics, and Business Fellows' Discussion Paper Series* 85 (2019).
- [2] Randall Albury. 1983. *The Politics of Objectivity*. Deakin University Press.
- [3] Michelle Alexander. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press.
- [4] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. *ProPublica* (2016). <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [5] Ruairi Arrieta-Kenna. 2018. 'Abolish Prisons' Is the New 'Abolish ICE'. *Politico* (2018). <https://www.politico.com/magazine/story/2018/08/15/abolish-prisons-is-the-new-abolish-ice-219361>
- [6] Amaryllis Austin. 2017. The Presumption for Detention Statute's Relationship to Release Rates. *Federal Probation* 81 (2017), 52.
- [7] David C. Baldus, Charles Pulaski, and George Woodworth. 1983. Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience. *Journal of Criminal Law and Criminology* 74, 3 (1983), 661–753.
- [8] Chelsea Barabas, Madars Virza, Karthik Dinakar, Joichi Ito, and Jonathan Zittrain. 2018. Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Vol. 81. PMLR, 62–76.
- [9] Shima Baradaran. 2011. Restoring the Presumption of Innocence. *Ohio State Law Journal* 72 (2011), 723–776.
- [10] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104 (2016), 671–732.
- [11] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. 2018. Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research* (2018), 1–42.
- [12] Robert Blizard. 2018. Key Findings from a National Survey of 800 Registered Voters January 11–14, 2018. (2018). <http://www.justiceactionnetwork.org/wp-content/uploads/2018/01/JAN-Poll-PPT-Jan25.2018.pdf>
- [13] James Bonta and Donald A. Andrews. 2007. Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation. *Public Safety Canada* (2007).
- [14] Meredith Broussard. 2018. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press.
- [15] Michelle Brown and Judah Schept. 2017. New abolition, criminology and a critical carceral studies. *Punishment & Society* 19, 4 (2017), 440–462.
- [16] Paul Butler. 2010. *Let's Get Free: A Hip-Hop Theory of Justice*. The New Press.
- [17] Paul Butler. 2016. The System Is Working the Way It Is Supposed to: The Limits of Criminal Justice Reform. *The Georgetown Law Journal* 104 (2016).
- [18] Paul Butler. 2017. *Chokehold: Policing Black Men*. The New Press.
- [19] Paul Butler. 2017. Equal Protection and White Supremacy. *Northwestern University Law Review* 112, 6 (2017), 1457–1464.
- [20] Paul D. Butler. 2012. Poor People Lose: Gideon and the Critique of Rights. *Yale Law Journal* 122 (2012), 2176–2204.
- [21] Samuel Carton, Jennifer Helsby, Kenneth Joseph, Ayesha Mahmud, Youngsoo Park, Joe Walsh, Crystal Cody, CPT Estella Patterson, Lauren Haynes, and Rayid Ghani. 2016. Identifying Police Officers at Risk of Adverse Events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, 67–76. <https://doi.org/10.1145/2939672.2939698>
- [22] Abram Chayes, William Fisher, Morton Horwitz, Frank Michelman, Martha Minow, Charles Nesson, and Todd Rakoff. [n.d.]. Critical Perspectives on Rights. ([n. d.]). <https://cyber.harvard.edu/bridge/CriticalTheory/rights.htm>
- [23] Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (2017), 153–163.
- [24] Julie Ciccolini and Cynthia Conti-Cook. 2018. Rationing Justice: Risk Assessment Instruments in the American Criminal Justice System. *EuropeNow* (2018). <https://www.europenowjournal.org/2018/11/07/rationing-justice-risk-assessment-instruments-in-the-american-criminal-justice-system/>
- [25] Rachel Cicurel. 2018. Motion to Exclude Results of the Violence Risk Assessment and All Related Testimony and/or Allocation Under FRE 702 and Daubert v. Merrell Dow Pharmaceuticals. (2018). <https://drive.google.com/open?id=1wA6VGpCA9-WVu48YYUksmLAIHs1utdi>
- [26] Patricia Hill Collins. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge.
- [27] Alexia Cooper and Erica L. Smith. 2011. Homicide Trends in the United States, 1980–2008. *U.S. Department of Justice, Bureau of Justice Statistics* (2011). <https://www.bjs.gov/content/pub/pdf/htus8008.pdf>
- [28] Sam Corbett-Davies, Sharad Goel, and Sandra González-Bailón. 2017. Even Imperfect Algorithms Can Improve the Criminal Justice System. *The New York Times* (2017). <https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html>
- [29] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 797–806. <https://doi.org/10.1145/3097983.3098095>
- [30] Robert M. Cover. 1986. Violence and the Word. *Yale Law Journal* 95 (1986), 1601–1629.
- [31] Bryce Covert. 2017. America Is Waking Up to the Injustice of Cash Bail. *The Nation* (2017). <https://www.thenation.com/article/america-is-waking-up-to-the-injustice-of-cash-bail/>
- [32] Bo Cowgill. 2018. The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities. (2018).
- [33] Kimberlé Williams Crenshaw. 1988. Race, Reform, and Retrenchment: Transformation and Legitimation in Antidiscrimination Law. *Harvard Law Review* 101, 7 (1988), 1331–1387.
- [34] Andrew Manuel Crespo. 2015. Systemic Facts: Toward Institutional Awareness in Criminal Courts. *Harvard Law Review* 129 (2015), 2049–2117.
- [35] Francis T. Cullen, Cheryl Lero Jonson, and Daniel S. Nagin. 2011. Prisons Do Not Reduce Recidivism: The High Cost of Ignoring Science. *The Prison Journal* 91, 3\_suppl (2011), 48S–65S. <https://doi.org/10.1177/0032885511415224>
- [36] Mona J.E. Danner, Marie VanNostrand, and Lisa M. Spruance. 2015. Risk-Based Pretrial Release Recommendation and Supervision Guidelines. *Luminosity, Inc.* (2015). <https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/risk-based-pretrial-release-recommendation-and-supervision-guidelines.pdf>
- [37] Lorraine Daston and Peter Galison. 2007. *Objectivity*. Zone Books.
- [38] Angela Y. Davis. 2003. *Are Prisons Obsolete?* Seven Stories Press.
- [39] Lois M. Davis, Robert Bozick, Jennifer L. Steele, Jessica Saunders, and Jeremy N.V. Miles. 2013. *Evaluating the Effectiveness of Correctional Education: A Meta-Analysis of Programs That Provide Education to Incarcerated Adults*. Rand Corporation.
- [40] Robert DeFina and Lance Hannon. 2010. For incapacitation, there is no time like the present: The lagged effects of prisoner reentry on property and violent crime rates. *Social Science Research* 39, 6 (2010), 1004–1014.
- [41] Richard Delgado and Jean Stefancic. 1992. Images of the Outsider in American Law and Culture: Can Free Expression Remedy Systemic Social Ills. *Cornell Law Review* 77, 6 (1992), 1258–1297.
- [42] Richard Delgado and Jean Stefancic. 2017. *Critical Race Theory: An Introduction* (third ed.). New York University Press.
- [43] William Dieterich, Christina Mendoza, and Tim Brennan. 2016. COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity. *Northpointe Inc. Research Department* (2016). [http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica\\_Commentary\\_Final\\_070616.pdf](http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf)
- [44] Will Dobbie, Jacob Goldin, and Crystal S. Yang. 2018. The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review* 108, 2 (2018), 201–40. <https://doi.org/10.1257/aer.20161503>
- [45] Laurel Eckhouse, Kristian Lum, Cynthia Conti-Cook, and Julie Ciccolini. 2019. Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment. *Criminal Justice and Behavior* 46, 2 (2019), 185–209.
- [46] EdBuild. 2019. \$23 Billion. (2019). <https://edbuild.org/content/23-billion/full-report.pdf>
- [47] Lee Fang. 2019. In Her First Race, Kamala Harris Campaigned as Tough on Crime – And Unseated the Country's Most Progressive Prosecutor. *The Intercept* (2019). <https://theintercept.com/2019/02/07/kamala-harris-san-francisco-district-attorney-crime/>
- [48] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and Removing Disparate Impact. In *21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 259–268. <https://doi.org/10.1145/2783258.2783311>
- [49] Joseph Fishkin. 2014. *Bottlenecks: A New Theory of Equal Opportunity*. Oxford University Press.
- [50] Data for Progress. 2018. Polling The Left Agenda. (2018). <https://www.dataforprogress.org/polling-the-left-agenda/>
- [51] Matt Ford. 2015. A New Approach to Criminal-Justice Reform. *The Atlantic* (2015). <https://www.theatlantic.com/politics/archive/2015/10/police-prosecutors-reform-group/411775/>
- [52] The Leadership Conference Education Fund. 2018. The Use of Pretrial "Risk Assessment" Instruments: A Shared Statement of Civil Rights Concerns. (2018). <https://leadershipconferenceedfund.org/pretrial-risk-assessment/>
- [53] FWD.us. 2018. Broad, Bipartisan Support for Bold Pre-Trial Reforms in New York State. (2018). <https://www.fwd.us/wp-content/uploads/2018/03/NYCJR-poll-memo-Final.pdf>
- [54] FWD.us. 2018. Every Second: The Impact of the Incarceration Crisis on America's Families. (2018). <https://everysecond.fwd.us/downloads/EverySecond.FWD.us.pdf>
- [55] Nazgol Ghandnoosh. 2014. Race and Punishment: Racial Perceptions of Crime and Support for Punitive Policies. *The Sentencing Project* (2014). <https://sentencingproject.org/wp-content/uploads/2015/11/Race-and-Punishment.pdf>
- [56] Phillip Atiba Goff, Matthew Christian Jackson, Di Leone, Brooke Allison Lewis, Carmen Marie Culotta, and Natalie Ann DiTomaso. 2014. The Essence of Innocence: Consequences of Dehumanizing Black Children. *Journal of Personality*



- and *Social Psychology* 106, 4 (2014), 526–545. <https://doi.org/10.1037/a0035663>
- [57] Shane Goldmacher. 2019. Michael Bloomberg Pushed 'Stop-and-Frisk' Policing. Now He's Apologizing. *The New York Times* (2019). <https://www.nytimes.com/2019/11/17/us/politics/michael-bloomberg-speech.html>
  - [58] Ben Green. 2018. Data Science as Political Action: Grounding Data Science in a Politics of Justice. *arXiv preprint arXiv:1811.03435* (2018).
  - [59] Ben Green. 2019. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*. MIT Press.
  - [60] Ben Green and Yiling Chen. 2019. Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* '19)*. ACM, 90–99. <https://doi.org/10.1145/3287560.3287563>
  - [61] Ben Green and Yiling Chen. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 50:1–50:24. <https://doi.org/10.1145/3359152>
  - [62] Ben Green, Thibaut Horel, and Andrew V. Papachristos. 2017. Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014. *JAMA Internal Medicine* 177, 3 (2017), 326–333. <https://doi.org/10.1001/jamainternmed.2016.8245>
  - [63] Ben Green and Lily Hu. 2018. The Myth in the Methodology: Towards a Re-contextualization of Fairness in Machine Learning. In *Machine Learning: The Debates workshop at the 35th International Conference on Machine Learning*.
  - [64] Kelly Hannah-Moffat, Paula Maurutto, and Sarah Turnbull. 2009. Negotiated Risk: Actuarial Illusions and Discretion in Probation. *Canadian Journal of Law & Society/La Revue Canadienne Droit et Société* 24, 3 (2009), 391–409.
  - [65] Lance Hannon and Robert DeFina. 2005. Violent Crime in African American and White Neighborhoods: Is Poverty's Detrimental Effect Race-Specific? *Journal of Poverty* 9, 3 (2005), 49–67. [https://doi.org/10.1300/J134v09n03\\_03](https://doi.org/10.1300/J134v09n03_03)
  - [66] Bernard E. Harcourt. 2015. Risk as a Proxy for Race: The Dangers of Risk Assessment. *Federal Sentencing Reporter* 27, 4 (2015), 237–243.
  - [67] Sandra Harding. 1998. *Is Science Multicultural?: Postcolonialisms, Feminisms, and Epistemologies*. Indiana University Press.
  - [68] Kamala Harris and Rand Paul. 2017. Pretrial Integrity and Safety Act of 2017. *115th Congress* (2017).
  - [69] Mark L. Hatzenbuehler, Katherine Keyes, Ava Hamilton, Monica Uddin, and Sandro Galea. 2015. The Collateral Damage of Mass Incarceration: Risk of Psychiatric Morbidity Among Nonincarcerated Residents of High-Incarceration Neighborhoods. *American Journal of Public Health* 105, 1 (2015), 138–143.
  - [70] Paul Heaton, Sandra Mayson, and Megan Stevenson. 2017. The downstream consequences of misdemeanor pretrial detention. *Stanford Law Review* 69 (2017), 711–794.
  - [71] Sara B. Heller. 2014. Summer jobs reduce violence among disadvantaged youth. *Science* 346, 6214 (2014), 1219–1223. <https://doi.org/10.1126/science.1257809>
  - [72] Deborah Hellman. 2019. Measuring Algorithmic Fairness. *Virginia Law Review* (2019).
  - [73] Elizabeth Kai Hinton, LeShae Henderson, and Cindy Reed. 2018. An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. (2018). <https://www.vera.org/publications/for-the-record-unjust-burden>
  - [74] Anna Lauren Hoffmann. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22, 7 (2019), 900–915. <https://doi.org/10.1080/1369118X.2019.1573912>
  - [75] Elie Honig. 2017. Elie Honig to Judge Grant. (2017). <https://assets.documentcloud.org/documents/3676485/AG-letter-asking-for-bail-reform-changes.pdf>
  - [76] Tom Jackman. 2016. U.S. police chiefs group apologizes for 'historical mistreatment' of minorities. *The Washington Post* (2016). <https://www.washingtonpost.com/news/true-crime/wp/2016/10/17/head-of-u-s-police-chiefs-apologizes-for-historic-mistreatment-of-minorities/>
  - [77] Sheila Jasanoff. 1986. *Risk Management and Political Culture*. Russell Sage Foundation.
  - [78] Sheila Jasanoff. 2004. The idiom of co-production. In *States of Knowledge: The Co-Production of Science and the Social Order*, Sheila Jasanoff (Ed.). Routledge, Chapter 1, 1–12.
  - [79] Sheila Jasanoff. 2004. Ordering knowledge, ordering society. In *States of Knowledge: The Co-Production of Science and the Social Order*, Sheila Jasanoff (Ed.). Routledge, Chapter 2, 13–45.
  - [80] Sheila Jasanoff. 2011. *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton University Press.
  - [81] Sheila Jasanoff. 2015. Future Imperfect: Science, Technology, and the Imaginations of Modernity. In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Sheila Jasanoff and Sang-Hyun Kim (Eds.). University of Chicago Press, Chapter 1, 1–47.
  - [82] Tom Jensen and John Tilley. 2012. HB 463 – Statement from the Sponsors. *Criminal Law Reform: The First Year of HB 463* (2012). [https://cdn.ymaws.com/www.kybar.org/resource/resmgr/2012\\_Convention\\_Files/ac2012\\_2.pdf](https://cdn.ymaws.com/www.kybar.org/resource/resmgr/2012_Convention_Files/ac2012_2.pdf)
  - [83] Jonathan Kahn. 2017. *Race on the Brain: What Implicit Bias Gets Wrong About the Struggle for Racial Justice*. Columbia University Press.
  - [84] Cheryl R. Kaiser, Brenda Major, Ines Jurcevic, Tessa L. Dover, Laura M. Brady, and Jenessa R. Shapiro. 2013. Presumed Fair: Ironic Effects of Organizational Diversity Structures. *Journal of Personality and Social Psychology* 104, 3 (2013), 504–519.
  - [85] Alec Karakatsanis. 2015. Policing, Mass Imprisonment, and the Failure of American Lawyers. *Harvard Law Review Forum* 128 (2015), 253.
  - [86] Alec Karakatsanis. 2019. The Punishment Bureaucracy: How to Think About "Criminal Justice Reform". *The Yale Law Journal Forum* 128 (2019), 848–935.
  - [87] Danielle Kehl, Priscilla Guo, and Samuel Kessler. 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. *Responsive Communities Initiative, Berkman Klein Center for Internet & Society* (2017).
  - [88] Duncan Kennedy. 2002. The Critique of Rights in Critical Legal Studies. In *Left Legalism/Left Critique*, Wendy Brown and Janet Halley (Eds.). Duke University Press, 178–228.
  - [89] Soohan Kim, Alexandra Kalev, and Frank Dobbin. 2012. Progressive Corporations at Work: The Case of Diversity Programs. *NYU Review of Law and Social Change* 36 (2012), 171.
  - [90] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2019. Discrimination in the Age of Algorithms. *Journal of Legal Analysis* 10 (2019), 113–174. <https://doi.org/10.1093/jla/laz001>
  - [91] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2016. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807* (2016).
  - [92] John Logan Koepke and David G. Robinson. 2018. Danger Ahead: Risk Assessment and the Future of Bail Reform. *Washington Law Review* 93 (2018), 1725–1807.
  - [93] Lauren J. Krivo, Ruth D. Peterson, and Danielle C. Kuhl. 2009. Segregation, Racial Structure, and Neighborhood Violent Crime. *Amer. J. Sociology* 114, 6 (2009), 1765–1802. <https://doi.org/10.1086/597285>
  - [94] Rachel Kushner. 2019. Is Prison Necessary? Ruth Wilson Gilmore Might Change Your Mind. *The New York Times* (2019). <https://www.nytimes.com/2019/04/17/magazine/prison-abolition-ruth-wilson-gilmore.html>
  - [95] Laura and John Arnold Foundation. 2018. Guide to the Release Conditions Matrix. (2018). [https://www.psapretrial.org/system/guides/guide\\_pdfs/000/000/012/original/9\\_Release\\_Conditions\\_Matrix.pdf?1529956534](https://www.psapretrial.org/system/guides/guide_pdfs/000/000/012/original/9_Release_Conditions_Matrix.pdf?1529956534)
  - [96] Laura and John Arnold Foundation. 2019. Public Safety Assessment (PSA) - Intro. (2019). <https://www.psapretrial.org/about>
  - [97] Lance Lochner and Enrico Moretti. 2004. The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review* 94, 1 (2004), 155–189. <https://doi.org/doi:10.1257/000282804322970751>
  - [98] German Lopez. 2019. Amy Klobuchar's record as a "tough on crime" prosecutor, explained. *Vox* (2019). <https://www.vox.com/policy-and-politics/2019/2/25/18225011/amy-klobuchar-president-prosecutor-criminal-justice-record>
  - [99] Christopher T. Lowenkamp and Jay Whetzel. 2009. The Development of an Actuarial Risk Assessment Instrument for U.S. Pretrial Services. *Federal Probation* 73 (2009).
  - [100] Kristian Lum, Erwin Ma, and Mike Baiocchi. 2017. The causal impact of bail on case outcomes for indigent defendants in New York City. *Observational Studies* 3 (2017), 39–64.
  - [101] Mona Lynch. 2016. *Hard Bargains: The Coercive Power of Drug Laws in Federal Court*. Russell Sage Foundation.
  - [102] Catharine A. MacKinnon. 1982. Feminism, Marxism, Method, and the State: An Agenda for Theory. *Signs: Journal of Women in Culture and Society* 7, 3 (1982), 515–544.
  - [103] Catharine A. MacKinnon. 2011. Substantive Equality: A Perspective. *Minnesota Law Review* 96 (2011).
  - [104] Martin Mahony. 2014. The predictive state: Science, territory and the future of the Indian climate. *Social Studies of Science* 44, 1 (2014), 109–133.
  - [105] Frank Main. 2016. Cook County judges not following bail recommendations: study. *Chicago Sun-Times* (2016). <https://chicago.suntimes.com/chicago-news/cook-county-judges-not-following-bail-recommendations-study-find/>
  - [106] Paula Maurutto and Kelly Hannah-Moffat. 2007. Understanding risk in the context of the Youth Criminal Justice Act. *Canadian Journal of Criminology and Criminal Justice* 49, 4 (2007), 465–491.
  - [107] Sandra G. Mayson. 2018. Dangerous Defendants. *Yale Law Journal* 127, 3 (2018), 490–568.
  - [108] Sandra G. Mayson. 2019. Bias In, Bias Out. *Yale Law Journal* 128, 8 (2019), 2218–2300.
  - [109] Allegra M. McLeod. 2013. Confronting Criminal Law's Violence: The Possibilities of Unfinished Alternatives. *Unbound: Harvard Journal of the Legal Left* 8 (2013), 109–132.
  - [110] Allegra M. McLeod. 2015. Prison Abolition and Grounded Justice. *UCLA Law Review* 62 (2015), 1156–1239.
  - [111] Allegra M. McLeod. 2019. Envisioning Abolition Democracy. *Harvard Law Review* 132 (2019), 1613–1649.
  - [112] Anne Milgram. 2014. Why smart statistics are the key to fighting crime. *TED* (2014). [https://www.ted.com/talks/anne\\_milgram\\_why\\_smart\\_statistics\\_are\\_](https://www.ted.com/talks/anne_milgram_why_smart_statistics_are_)

- the\_key\_to\_fighting\_crime/transcript?language=en
- [113] Alex P. Miller. 2018. Want Less-Biased Decisions? Use Algorithms. *Harvard Business Review* (2018). <https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms>
  - [114] Martha Minow. 1991. *Making All the Difference: Inclusion, Exclusion, and American Law*. Cornell University Press.
  - [115] John Monahan and Jennifer L. Skeem. 2016. Risk Assessment in Criminal Sentencing. *Annual Review of Clinical Psychology* 12 (2016), 489–513.
  - [116] Evgeny Morozov. 2014. *To Save Everything, Click Here: The Folly of Technological Solutionism*. PublicAffairs.
  - [117] Samuel Moyn. 2015. Civil Liberties and Endless War. *Dissent* (2015). <https://www.dissentmagazine.org/article/civil-liberties-and-endless-war>
  - [118] Naomi Murakawa. 2014. *The First Civil Right: How Liberals Built Prison America*. Oxford University Press.
  - [119] La Vonne I. Neal, Audrey Davis McCray, Gwendolyn Webb-Johnson, and Scott T. Bridgest. 2003. The Effects of African American Movement Styles on Teachers' Perceptions and Reactions. *The Journal of Special Education* 37, 1 (2003), 49–57. <https://doi.org/10.1177/00224669030370010501>
  - [120] New Jersey Courts. 2017. One Year Criminal Justice Reform Report to the Governor and the Legislature. (2017). <https://www.njcourts.gov/courts/assets/criminal/2017cjrannual.pdf>
  - [121] Northpointe, Inc. 2015. Practitioner's Guide to COMPAS Core. (2015). <http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-031915.pdf>
  - [122] Northpointe, Inc. 2016. Sample-COMPAS-Risk-Assessment-COMPAS-“CORE”. (2016). <https://assets.documentcloud.org/documents/2702103/Sample-Risk-Assessment-COMPAS-CORE.pdf>
  - [123] Cathy O'Neil. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.
  - [124] Devah Pager. 2003. The Mark of a Criminal Record. *Amer. J. Sociology* 108, 5 (2003), 937–975.
  - [125] Gary Peller. 1990. Race Consciousness. *Duke Law Journal* (1990), 758.
  - [126] Julie A. Phillips. 2014. White, Black, and Latino Homicide Rates: Why the Difference? *Social Problems* 49, 3 (2014), 349–373. <https://doi.org/10.1525/sp.2002.49.3.349>
  - [127] Christopher S. Porrino. 2017. Attorney General Law Enforcement Directive 2016-6 v3.0. (2017). [https://www.nj.gov/lps/dcj/agguide/directives/ag-directive-2016-6\\_v3-0.pdf](https://www.nj.gov/lps/dcj/agguide/directives/ag-directive-2016-6_v3-0.pdf)
  - [128] Christopher S. Porrino. 2017. Attorney General Law Enforcement Directive No. 2016-6 v2.0. (2017). [https://nj.gov/oag/newsreleases17/Revised-AG-Directive-2016-6\\_Introductory-Memo.pdf](https://nj.gov/oag/newsreleases17/Revised-AG-Directive-2016-6_Introductory-Memo.pdf)
  - [129] Theodore M. Porter. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.
  - [130] Pretrial Justice Institute. 2017. Pretrial Risk Assessment Can Produce Race-Neutral Results. (2017). <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.aspx?DocumentFileKey=5cebc2e7-dfa4-65b2-13cd-300b81a6ad7a>
  - [131] Pretrial Justice Institute. 2017. The State of Pretrial Justice in America. (2017). <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.aspx?DocumentFileKey=f9d452f6-ac5a-b8e7-5d68-0969abd2cc82&forceDialog=0>
  - [132] Pretrial Justice Institute. 2019. Scan of Pretrial Practices. *Pretrial Justice Institute* (2019). <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.aspx?DocumentFileKey=24bb2bc4-84ed-7324-929c-d0637db43c9a&forceDialog=0>
  - [133] Seth J. Prins and Adam Reich. 2017. Can we avoid reductionism in risk reduction? *Theoretical Criminology* (2017), 1–21.
  - [134] Gideon's Promise, The National Legal Aid, Defenders Association, The National Association for Public Defense, and The National Association of Criminal Defense Lawyers. 2017. Joint Statement in Support of the Use of Pretrial Risk Assessment Instruments. (2017). <http://www.publicdefenders.us/files/Defenders%20Statement%20on%20Pretrial%20RAI%20May%202017.pdf>
  - [135] Lincoln Quillian and Devah Pager. 2001. Black neighbors, higher crime? The role of racial stereotypes in evaluations of neighborhood crime. *Amer. J. Sociology* 107, 3 (2001), 717–767.
  - [136] Marvin Rausand. 2013. *Risk Assessment: Theory, Methods, and Applications*. John Wiley & Sons.
  - [137] Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability. (2018). <https://ainowinstitute.org/aiareport2018.pdf>
  - [138] Harvard Law Review. 2019. Introduction. *Harvard Law Review* 132, 6 (2019), 1568–1574.
  - [139] Dina R. Rose and Todd R. Clear. 1998. Incarceration, Social Capital, and Crime: Implications for Social Disorganization Theory. *Criminology* 36, 3 (1998), 441–480.
  - [140] Mica Rosenberg and Reade Levinson. 2018. Trump's catch-and-detain policy snares many who have long called U.S. home. *Reuters* (2018). <https://www.reuters.com/investigates/special-report/usa-immigration-court/>
  - [141] Richard Rothstein. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright Publishing Corporation.
  - [142] Robert J. Sampson. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. University of Chicago Press.
  - [143] Robert J. Sampson, Jeffrey D. Morenoff, and Stephen Raudenbush. 2005. Social Anatomy of Racial and Ethnic Disparities in Violence. *American Journal of Public Health* 95, 2 (2005), 224–232. <https://doi.org/10.2105/AJPH.2004.037705>
  - [144] Robert J. Sampson and William Julius Wilson. 1995. Toward a Theory of Race, Crime, and Urban Inequality. In *Crime and Inequality*, John Hagan and Ruth Peterson (Eds.). Vol. 1995. 37–54.
  - [145] Jon Schuppe. 2017. Post Bail. *NBC News* (2017). <https://www.nbcnews.com/specials/bail-reform>
  - [146] Alexander Shalom, Colette Tvedt, Joseph E. Krakora, and Diane DePietropaolo Price. 2016. The New Jersey Pretrial Justice Manual. (2016). <http://www.nacdl.org/njpretrial>
  - [147] Jennifer L. Skeem and Christopher T. Lowenkamp. 2016. Risk, Race, & Recidivism: Predictive Bias and Disparate Impact. *Criminology* 54, 4 (2016), 680–712. <https://doi.org/10.1111/1745-9125.12123>
  - [148] Sonja B. Starr. 2014. Evidence-Based Sentencing and the Scientific Rationalization of Discrimination. *Stanford Law Review* 66, 4 (2014), 803–872.
  - [149] David Steinhart. 2006. Juvenile detention risk assessment: A practice guide to juvenile detention reform. *The Annie E. Casey Foundation* (2006). <https://www.aecf.org/m/resourceimg/aecf-juvenile-detention-risk-assessment1-2006.pdf>
  - [150] Bryan Stevenson. 2019. Why American Prisons Owe Their Cruelty to Slavery. *The New York Times Magazine* (2019). <https://www.nytimes.com/interactive/2019/08/14/magazine/prison-industrial-complex-slavery-racism.html>
  - [151] Megan T. Stevenson. 2017. Risk Assessment: The Devil's in the Details. *The Crime Report* (2017). <https://thecrimereport.org/2017/08/31/does-risk-assessment-work-theres-no-single-answer/>
  - [152] Megan T. Stevenson. 2018. Assessing Risk Assessment in Action. *Minnesota Law Review* 103 (2018).
  - [153] Megan T. Stevenson and Jennifer L. Doleac. 2018. The Roadblock to Reform. *The American Constitution Society* (2018). <https://www.acslaw.org/wp-content/uploads/2018/11/RoadblockToReformReport.pdf>
  - [154] Cass R. Sunstein. 2018. Algorithms, Correcting Biases. *Social Research* (2018).
  - [155] Jared Sylvester and Edward Raff. 2018. What About Applied Fairness? In *Machine Learning: The Debates Workshop at the 35th International Conference on Machine Learning*.
  - [156] Mark Tushnet. 1983. An Essay on Rights. *Texas Law Review* 62, 8 (1983), 1363–1403.
  - [157] Mark Tushnet. 1993. The Critique of Rights. *SMU Law Review* 47 (1993), 23–34.
  - [158] Cody Tuttle. 2019. Snapping Back: Food Stamp Bans and Criminal Recidivism. *American Economic Journal: Economic Policy* 11, 2 (2019), 301–27. <https://doi.org/10.1257/pol.20170490>
  - [159] United States Department of Justice Federal Bureau of Investigation. 2018. Murder Offenders by Age, Sex, Race, and Ethnicity, 2017. *Crime in the United States* (2018). <https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/tables/expanded-homicide-data-table-3.xls>
  - [160] United States Sentencing Commission. 1987. Sentencing Guidelines and Policy Statements. (1987).
  - [161] U.S. Supreme Court. 1987. *McCleskey v. Kemp*. 481 U.S. 279.
  - [162] U.S. Supreme Court. 1987. *United States v. Salerno*. 481 U.S. 739.
  - [163] Arnold Ventures. 2019. Public Safety Assessment FAQs (“PSA 101”). (2019). [https://craftmediabucket.s3.amazonaws.com/uploads/Public-Safety-Assessment-101\\_190319\\_140124.pdf](https://craftmediabucket.s3.amazonaws.com/uploads/Public-Safety-Assessment-101_190319_140124.pdf)
  - [164] Arnold Ventures. 2019. Statement of Principles on Pretrial Justice and Use of Pretrial Risk Assessment. (2019). <https://craftmediabucket.s3.amazonaws.com/uploads/Arnold-Ventures-Statement-of-Principles-on-Pretrial-Justice.pdf>
  - [165] Lynne M. Vieraitis, Tomislav V. Kovandzic, and Thomas B. Marvell. 2007. The Criminogenic Effects of Imprisonment: Evidence from State Panel Data, 1974–2002. *Criminology & Public Policy* 6, 3 (2007), 589–622.
  - [166] Shiv Visvanathan. 2005. Knowledge, justice and democracy. In *Science and Citizens: Globalization and the Challenge of Engagement*, Melissa Leach, Ian Scoones, and Brian Wynne (Eds.). Zed Books.
  - [167] Human Rights Watch. 2017. “Not in it for Justice”: How California's Pretrial Detention and Bail System Unfairly Punishes Poor People. (2017). <https://www.hrw.org/report/2017/04/11/not-it-justice/how-californias-pretrial-detention-and-bail-system-unfairly>
  - [168] Bruce Western. 2006. *Punishment and Inequality in America*. Russell Sage Foundation.
  - [169] Rebecca Wexler. 2017. Code of Silence. *Washington Monthly* (2017). <https://washingtonmonthly.com/magazine/junejulyaugust-2017/code-of-silence/>
  - [170] Christopher Wildeman and Emily A. Wang. 2017. Mass incarceration, public health, and widening inequality in the USA. *The Lancet* 389, 10077 (2017), 1464–1474.
  - [171] Wisconsin Supreme Court. 2016. *State v. Loomis*. 881 Wis. N.W.2d 749.
  - [172] Crystal S. Yang. 2017. Toward an Optimal Bail System. *New York University Law Review* 92, 5 (2017), 1399–1493.