Affirmative Algorithms: Relational Equality as Algorithmic Fairness

Marilyn Zhang
zmarilyn@cs.stanford.edu
Stanford University
Stanford, CA, USA

ABSTRACT

Many statistical fairness notions have been proposed for algorithmic decision-making systems, and especially public safety pretrial risk assessment (PSPRA) algorithms such as COMPAS. Most fairness notions equalize something between groups, whether it is false positive rates or accuracy. In fact, I demonstrate that most prominent notions have their basis in equalizing some form of accuracy. However, statistical fairness metrics often do not capture the substantive point of equality. I argue that equal accuracy is not only difficult to measure but also unsatisfactory for ensuring equal justice. In response, I introduce philosopher Elizabeth Anderson's theory of relational equality as a fruitful alternative framework: to relate as equals, people need access to certain basic capabilities. I show that relational equality requires Affirmative PSPRA algorithms that lower risk scores for Black defendants. This is because fairness based on relational equality means considering the impact of PSPRA algorithms' decisions on access to basic capabilities. This impact is racially asymmetric in an unjust society. I make three main contributions: (1) I illustrate the shortcomings of statistical fairness notions in their reliance on equalizing some form of accuracy; (2) I present the first comprehensive ethical defense of Affirmative PSPRA algorithms, based on fairness in terms of relational equality instead; and (3) I show that equalizing accuracy is neither sufficient nor necessary for fairness based on relational equality. Overall, this work serves narrowly as a reason to re-evaluate algorithmic fairness for PSPRA algorithms, and serves broadly as an example of how discussions of algorithmic fairness can benefit from egalitarian philosophical frameworks.

CCS CONCEPTS

• **Social and professional topics** → **Computing / technology policy**; • **Computing methodologies** → *Artificial intelligence*; • **Applied computing** → *Law, social and behavioral sciences.*

KEYWORDS

fairness, algorithmic fairness, philosophy, relational equality, affirmative algorithms, criminal justice, pretrial risk assessments

1 INTRODUCTION

Public safety pretrial risk assessment (PSPRA) algorithms attempt to predict whether a defendant poses a public safety risk before their trial. Many existing algorithms, however, have been criticized for being unfair and biased against Black defendants (e.g. [8, 37, 58]). In response, fairness metrics for algorithmic decision-making systems such as PSPRA algorithms have been proposed and widely discussed (e.g. [17, 26, 39, 100]). Most of these fairness notions aim to equalize something between groups, whether it is false positive rates or accuracy (e.g. [27, 47]; see also Table 1).

The problem with statistical equality, however, is that setting it as the goal for algorithmic fairness risks "treat[ing] everyone the same from an algorithmic perspective without acknowledging that people are not treated the same" [45, 9]. In an unjust society, the same decision can have asymmetric consequences for different groups of people.[1] Hence there has been a call to understand algorithmic fairness beyond statistical equality by drawing from other fields such as philosophy, feminist studies, and critical race theory [18, 43, 45]. More broadly, there has been a push to understand algorithms and fairness within real-world applications and conditions instead of in isolation (e.g. [14, 34, 44, 111]).

This paper continues the conversation by drawing from a philosophical conception of equality — relational equality [5] — as a desirable alternative to statistical equality in PSPRA algorithms. Relational equality stipulates that equality has been achieved when everyone is able to relate as moral equals and participate fully in a democracy [5, 315]. Rather than focusing on material equality, relational equality is primarily concerned with social relations [5, 312]. As such, it is especially apt for considering racial fairness in PSPRA algorithms, which is the subject of this paper.

Assuming that relational equality really does capture what matters in equality, I argue that Affirmative PSPRA algorithms would be required from a fairness perspective. Risk scores for Black defendants would need to be lowered. Throughout, I continue the standard of focusing on white and Black defendants for racial fairness in PSPRA algorithms (e.g. [8, 28]). Furthermore, I do not argue for or against using PSPRA algorithms in general; in fact, perhaps abolishing pretrial detention may be the most fair (e.g. [42, 71, 86]).

---

[1][34, 60] makes a similar critique that current fairness ideals disregard how the same burden can affect different people differently.

Rather, I argue that where PSPRA algorithms are used, an Affirmative PSPRA algorithm would be the most racially fair version based on relational equality.

I begin in §2 with background context on the appeal of PSPRA algorithms and the dominance of equal accuracy as the foundation for understanding their racial fairness. In §3 and §4, I demonstrate the shortcomings of equalizing accuracy as the primary basis for fairness. §3 shows that accuracy is problematic to measure, and §4 argues that equalizing accuracy does not necessarily entail equal justice. In response to these shortcomings of equalizing accuracy, §5 provides an alternative framework for equality: philosopher Elizabeth Anderson's theory of relational equality.[20] The implications of fairness based on relational equality instead are discussed in §6. Specifically, Affirmative PSPRA algorithms that lower risk scores for Black defendants would be the most fair version of a PSPRA algorithm. An unjust society means PSPRA algorithms' decisions can be especially harmful for Black communities' access to basic capabilities necessary for relational equality. Affirmative PSPRA algorithms also lead to the conclusion that equalizing accuracy is neither sufficient nor necessary for fairness based on relational equality.

In summary, I demonstrate the shortcomings of statistical fairness notions, and respond by providing the first comprehensive perspective from egalitarian philosophy[19] as to what fairness in PSPRA algorithms requires. This undertaking also doubles as the first ethical defense of Affirmative PSPRA algorithms.[23] Finally, the philosophical steps applied to PSPRA algorithms in this paper serves as a model for how political philosophy can concretely weigh in on algorithmic fairness questions in other contexts in the future.

2 THE DOMINANCE OF EQUAL ACCURACY IN ALGORITHMIC FAIRNESS

2.1 The Appeal of PSPRA Algorithms

PSPRA algorithms attempt to predict a defendant's public safety risk in an objective manner [25]. PSPRA algorithms are used by judges after a defendant is arrested, but before their trial, to help determine whether a defendant should be incarcerated before their trial [25]. The inputs to a PSPRA algorithm is a defendant's profile, which can include a defendant's age, prior arrests, and gender.[2] The output is a risk score, for example from 1 to 10 [32, 31], of whether a defendant will be re-arrested for another crime if they are released before their trial. Examples of PSPRA algorithms include Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) developed by Equivant [32] and Public Safety Assessment (PSA) developed by Arnold Ventures [9].[3] PSPRA algorithms have been deployed widely across the United States.[4]

What is the appeal of PSPRA algorithms? For one, PSPRA algorithms are seen as the primary alternative to a broken cash bail

system, short of abolishing pretrial detention altogether.[5] In replacing the cash bail system so that defendants are not detained simply because they cannot afford to pay bail, PSPRA algorithms are also seen as tools that can help decrease the current pretrial detention rate with few public safety risks [65].[6] Second, by serving as a reference for human judges, risk scores from PSPRA algorithms could speed up the decision-making process for detaining or releasing defendants [53]. Finally, using PSPRA algorithms could potentially avoid racial bias (by excluding race as an input) [98, 508], as well as other cognitive biases [98, 501-2], when assessing risk.

2.2 The Dominance of Equal Accuracy in Algorithmic Fairness

Perhaps even more so than the promise of PSPRA algorithms, the perils of PSPRA algorithms have been widely discussed (e.g. [8, 37, 95]). The primary concern is that PSPRA algorithms are racially unfair. But how do we define racial fairness? Many fairness notions have arisen in response. Table 1 includes some of the most prominent fairness notions, although too many exist to pursue a comprehensive survey here.[7] The bulk of prominent fairness notions are based on equalizing a statistical metric of some sort – in fact, all in Table 1 are. It is important to note, however, that there are some other prominent fairness notions that are not. Chief among them is anti-classification [27, 5], which is a form of procedural fairness that recommends being race-blind during the decision-making process.[8] There is also fairness through awareness [30], which typically provides a process for fairness.

The following are definitions for terms referenced in Table 1 and throughout this paper. **Score** is the PSPRA algorithm's predicted risk score for a particular defendant; this can be binarized into a positive (will recidivate) or negative (will not recidivate) prediction. **Outcome** is whether the defendant actually would have recidivated had they been released pretrial. **False positives** detain defendants who would not have committed a crime had they been released. **False negatives** release defendants who do commit a crime when they are released. **False positive rate** is the proportion of detained defendants who would not have committed a crime had they been released. **True positive rate** is the proportion of detained defendants who would have committed a crime had they been released. **False negative rate** is the proportion of released defendants who do commit a crime. **Accuracy** is the proportion of correct predictions out of all the predictions made by a PSPRA algorithm. Less formally, accuracy has different variations, which are further clarified in Table 1.

Moreover, most of these statistical equality notions are based not only in equality, but in fact on some notion of equal accuracy. This is also depicted in Table 1. The shared reliance of these statistical

---

[2]Specific inputs depend on the particular PSPRA algorithm. For an example of inputs, see [7, 9].
[3]Unlike COMPAS, PSA does not take into account factors like education, income, or employment, which are more correlated with certain demographics.
[4]PSA alone is used in 22 different states, in some instances across the entire state and specific counties for others, with validation studies conducted for each [23]. In California, 49 out of 58 counties currently use PSPRA algorithms [48].

[5]In fact, two formerly incarcerated men wanting to rehaul the cash bail system penned the precursor bill to CA Prop 25, which proposes to use PSPRA algorithms instead of cash bail [62].
[6]One of the reasons Arnold Ventures developed PSA was to lower the pretrial detention rate [9], and New Jersey was able to reduce its pretrial population by 50% between 2015-18 by shifting to PSPRA algorithms [55]. However, [101] found PSPRA algorithms to have little effect on lowering pretrial detention.
[7]For reviews of fairness notions, see [17, 26, 39, 100]. [100, 2] also includes citation counts as of 2018 for the papers that introduced each fairness notion. Inclusion as a prominent fairness notion in Table 1 was based on these citation counts.
[8]In some contexts, anti-classification is legally required. Whether this is true in the PSPRA algorithm case is unclear. See for example [51, 56, 108].

Table 1: Most prominent statistical fairness notions equalize some variation of accuracy.

Statistical Fairness Notion	Description	Equalizes Some Variation of Accuracy?
Balance [66]	Average predicted score for all Black defendants with a certain outcome must be equal to the average predicted score for all white defendants with the same outcome [66, 4].	**Yes.** Ensures for defendants with the same outcome that scores aren't "systematically more inaccurate" [66, 4] for one racial group.
Calibration [66]	For defendants given a specific score, the probability of recidivating should be equal across racial groups. [66, 4].	**Yes.** Equalizes predictive accuracy for racial groups conditional on score.
Classification Parity [27]	The false positive rate should be equal across racial groups, as with false negative rates. [27, 6].	**Yes.** Equalizes error rates, which are complements to accuracy (informally).
Equalized Odds [47]	The false positive rate should be equal across racial groups, as with true positive rates [47, 2-3].	**Yes.** Enforces "equal bias and equal accuracy" [47, 3] across racial groups.
Predictive Parity [66]	For defendants given a positive prediction, the probability of recidivating should be equal across racial groups. [100, 3].	**Yes.** Similar to calibration, but uses a binarized version of score instead [100, 5].
Statistical Parity [30]	The probability of white defendants and Black defendants receiving a positive (or negative) prediction should be equal [100, 3].	**No.** Equalizes an outcome but not accuracy.

equality notions on accuracy (and equalizing accuracy) reflects a traditional primacy of accuracy as a metric and value in PSPRA algorithms: the higher the accuracy at predicting whether someone will recidivate before their trial, the better. And in many instances, maximizing accuracy is a reasonable primary goal. Let Algorithm 1 be a PSPRA algorithm that does not take into account age. Algorithm 1 achieves 50% accuracy for individuals aged 65 and over, as well as 50% accuracy for individuals 25 and under. By taking into account age, the accuracy improves to 70% for both age groups.[9] Let us call this Algorithm 2. In this scenario, it seems that Algorithm 2 is more desirable than Algorithm 1, because it has taken into account a confounding variable — age — that increases its accuracy. Consider a 20 year old and an 80 year old with the same arrest record, who both get labeled as low risk by Algorithm 1. Not including age unfairly penalizes the older individual, who among many other factors has had more time to rack up the same arrest record. This is why the higher accuracy in Algorithm 2 achieved by including age as an input variable is desirable.

For these reasons, accuracy is often considered not only a desirable quality of an algorithm but the most desirable quality — in fact, it is the "most common" value found in top computer science papers [19, 5]. Even outside of computer science communities, algorithms tend to be associated with maximizing accuracy. For example, legal scholar Sandra Mayson argues that having a single risk threshold for all defendants is a "sine qua non of risk assessment" precisely because the job of these algorithms is to label high risk defendants as high risk and low risk defendants as low risk [74, 2274]. Thus it is almost natural that when we intuitively conceive of fairness as equalizing something between groups, some variation of accuracy is equalized.

3 PROBLEMS WITH MEASURING ACCURACY

The next two sections criticize the current dominance of equalizing accuracy as the foundation for fairness notions. Although I focus on equalizing accuracy as the subject, these critiques extend to the primacy of accuracy as well. Furthermore, although the different fairness notions presented earlier are based on equalizing different variations of accuracy, the criticisms presented do not rely on these nuances. Rather, this section underscores fundamental issues with *measuring* any type of accuracy in PSPRA algorithms (which is necessary to equalize accuracy). §4 then questions the *value* of equalizing accuracy: it does not necessarily translate to equal justice. These discussions lead to the ultimate conclusion that equalizing accuracy in any form, contrary to common understanding, is unsatisfactory for fairness.

3.1 Measuring Recidivism

Accuracy is an unreliable metric in PSPRA algorithms because measuring recidivism can introduce bias.

The pipeline to measure accuracy in PSPRA algorithms is roughly as follows. First, an algorithm predicts whether a given defendant is high risk for recidivating or low risk. Then, this risk score is compared with what actually happens to the defendant. Cases where the defendant was predicted to be high risk and actually recidivated, or if the defendant was predicted to be low risk and did not recidivate, are considered as correct predictions.[10] Other scenarios are considered as wrong predictions. The proportion of all predictions that are correct is the PSPRA algorithm's accuracy. The problem happens in the second step. In order to understand whether

---

[9]This is not an unreasonable scenario, given the correlation between age and recidivism. See, for example [59].

[10]It is difficult to measure what happens in the counterfactual, namely if a defendant is detained pretrial there is no way of knowing whether the defendant would have recidivated or not had they been released pretrial. Hence, most studies that calculate accuracy make a retroactive comparison by calculating risk scores only for defendants who have been released [27, 18].

a prediction was correct or wrong, it must be known whether the defendant actually recidivated or not. But measuring recidivism can introduce bias, especially when recidivism is defined as a case of re-arrest – a poor proxy [13, 3]. Racial bias in police practices[11] likely inflates arrest rates for Black individuals [70, 78, 82], meaning measured recidivism rates (compared to actual recidivism rates) are likely disproportionately higher for Black defendants than white defendants.

The upshot is that measuring accuracy depends heavily on measuring recidivism, and measuring recidivism likely reflects racial bias in policing practices. This makes it difficult to guarantee that the number being output as "accuracy" is untainted. Hence, even if we value equalizing accuracy, using accuracy as a metric for PSPRA algorithms can reflect bias because it is dependent on sociological factors like policing practices.

3.2 Short Time Horizon

Measuring accuracy in predicting pretrial recidivism is also problematic because it captures too short a time horizon to meet the goals of measuring accuracy in the first place.

What are the goals of PSPRA algorithms? Most narrowly, the goal of a PSPRA algorithm is to accurately predict whether a given defendant will recidivate before their trial. Why? It is generally accepted that recidivism impacts public safety, and public safety is important. But in that case, why not just incarcerate everyone pretrial? Individual liberty is also important. Hence accuracy is supposed to maximize two important values: public safety and individual liberty. The broader goal of a PSPRA algorithm is to incarcerate as few individuals as possible while minimizing public safety costs.

Maximizing public safety and individual liberty are both desirable goals to have. However, perhaps somewhat counterintuitively, accurately predicting whether someone will recidivate before their trial does not always best serve these goals. PSPRA algorithms traditionally predict whether someone will recidivate before their trial. This is taken as the only source of public safety risk.

But this time horizon is too short. If what we really care about is public safety, we must also consider whether someone will recidivate after their trial. This is relevant in the pretrial stage because falsely incarcerating someone pretrial, who would not have otherwise recidivated, renders them more likely to recidivate after their trial. For example, a 2013 study from the Laura and John Arnold Foundation found that in contrast to releasing low-and moderate-risk defendants, "detaining low-and moderate-risk defendants,[12] even just for a few days, is strongly correlated with higher rates of new criminal activity both during the pretrial period and years after case disposition" [69, 3]. A 2018 study that observed defendants for two full years post-arrest corroborated these findings: even after controlling for baseline recidivism risk, defendants detained for at least three days are more likely to be rearrested compared to defendants released within three days [29, 226-7]. Hence if what

we care about is public safety, then the scope of recidivism must be broadened to consider more than pretrial recidivism.[13]

The upshot is that if PSPRA algorithms are meant to maximize both public safety and individual liberty, they cannot merely accurately predict whether someone will recidivate before their trial. It is certainly easier to consider the direct effects of false negatives on public safety, where someone recidivates *before* their trial. But crucially, false positives are also implicated in public safety, where someone is more likely to recidivate *after* their trial. False positives also restrict individual liberty by incarcerating someone before their trial. Hence to maximize both public safety and individual liberty, it is not enough to maximize accuracy in predicting pretrial recidivism alone. The time horizon must be expanded to incorporate post-trial effects.

Thus any notion of PSPRA algorithmic fairness that relies on accuracy, such as equalizing accuracy, is unsound; accuracy the way we measure it now does not capture what we want it to. It is a flawed metric.

4 RECONSIDERING THE VALUE OF EQUALIZING ACCURACY

Beyond *measurement* difficulties captured in §3.1 and §3.2, equalizing accuracy may also have questionable *value* in the first place: I show that if what we care about is fairness and justice, equal accuracy does not mean that equal justice has been served. This is partially because of a currently unjust basic structure. Hence, §5 and §6 propose shifting away from statistical notions of equality in favor of an alternative value (relational equality); I end this section by showing that there is room in the criminal justice system for such a value.

4.1 Unjust Basic Structure

Philosopher John Rawls defines the basic structure of society as the collection of major social and political institutions in society [85, 159]. These include the economy, the family, the legal system, and the political constitution [104]. The basic structure of society is what decides how the main benefits and burdens of social cooperation are distributed. This includes, for example, who receives social recognition and how income and wealth are distributed [104]. Social stigma impacts who receives social recognition.[14] Loan policies help determine who gets more chances for upward mobility by doing something meaningful with a loan. The quality of an education system can have an impact on what opportunities are available later in life, which also translates into income and wealth distributions. Social stigma, loan policies, and educational opportunities are all components of the basic structure of society.

---

[11]Besides current bias, even past racial bias in police practices can lead to overpolicing in certain areas today when prior data is fed into predictive policing algorithms. See [70].
[12]We can think of these as false positives.

[13]Note, however, that individual liberty, not incarceration, exists by default in a liberal democracy. There is an asymmetrically higher standard of justification for incarceration, especially pretrial incarceration, compared to pretrial release (after all, the defendant has not had a trial to be proven guilty yet and is therefore innocent). Hence while the causal relationship between pretrial incarceration and post-trial recidivism may justify releasing a defendant pretrial (when pretrial incarceration causes post-trial recidivism), it does not directly lead to the corollary of incarcerating a defendant pretrial (when pretrial incarceration reduces post-trial recidivism).
[14]Although Rawls does not focus on social stigma in his original formulation of the basic structure, Iris Marion Young argues that a "more direct evaluation" of the basic structure reveals it to include social stigma [110].

As philosopher Tommie Shelby has argued, the United States has a racially unjust basic structure today. A history of slavery followed by formal Jim Crow segregation against Black people led to unequal access to resources, schools, deep economic exploitation, and stereotypes of inferiority, among many other injustices [94, 43]. In the current system, structurally racist policies like discrimination in employment and exclusionary zoning and redlining reinforce the racially unjust basic structure [89, 90, 94]. New York City schools, for example, are still racially segregated due to housing policies and selective admissions procedures [46, 93]. Furthermore, the "stigma of race has become the stigma for criminality" [3, 248]. Being young, Black, and male is used to justify the "arrest, interrogation, search, and detention" of thousands of Black people every year, often taking the form of police brutality and violence [3, 248]. This often also leads to unwarranted exclusion from housing, employment, and educational opportunities [3, 248]. Beyond formal exclusion, racial stigma also leads to informal exclusion, manifested in who people choose to associate with [41, 71]. Racial stigma broadly "triggers. . . exclusion and invisibility" [64, 79]. Finally, technology is often complicit in contributing to the racially unjust basic structure: predictive policing algorithms such as PredPol uses past crime data to target predominantly Black neighborhoods, creating a self-fulfilling prophecy that Ruha Benjamin has termed them "crime production algorithms" [15, 82-3].

4.2 Equal Accuracy is not Equal Justice

Partially because of an unjust basic structure, equal accuracy in PSPRA algorithms does not entail equal justice.

A common intuition is that if a task X is desirable, then an algorithm having lower accuracy for task X for some group while having higher accuracy for task X for another is undesirable. For example, we should (and do) react poorly when IBM's face recognition technology correctly classifies lighter-skinned individuals 96.8% of the time but correctly classifies darker-skinned individuals only 77.6% of the time [21, 9]. Existing literature is also wary of the common side effect of maximizing overall accuracy: majority groups tend to get higher accuracy rates than minority groups [26, 2].

Yet while equalizing accuracy may seem intuitive, I show that equalizing accuracy is not sufficient for equal justice (and in fact, in §6 I show that it is not necessary, either). Specifically, I show that even if accuracy is equal between two groups, that does not necessarily mean equal justice has been served for the two groups. This is why equalizing accuracy between groups is a flawed value.

Why does equal accuracy not necessarily entail equal justice? Despite equal accuracy, especially harmful errors can be concentrated towards a particular group; I call this *concentrating injustice*. As Renée Jorgensen asserts, "if an evidence rule concentrates the risk of suffering false findings on a subgroup of the population . . . members of that group have a justice-based complaint against the rule proportional to the severity and concentration of the risk" [20, 70]. As an example, the COMPAS algorithm is equally accurate for both white and Black defendants, at roughly 60% accuracy [8]. Yet, the false positive rate is almost twice as high for Black defendants [8]. False positives are more costly than false negatives to individual defendants: being unfairly incarcerated is a loss of individual

liberty, whereas recidivating when released may not even have any costs to individual defendants if they are not arrested. Hence this disparity in false positive rates concentrates injustices on Black defendants more so than on white defendants.

In the previous example, the concentrated injustice occurred in the outcomes of false positives. But *concentrating* injustice can occur even if the bad outcome one is at risk for is not actually realized. Consider a surveillance system that is equally accurate for Black and white people, but has a much higher false positive rate (for being marked a criminal) for Black people than white people. On top of the harms suffered *after* a false positive occurs, even *before* a false positive happens, members of Black communities can suffer psychological harm and restrict their movements to protect themselves from being targeted by an unfair scheme [20, 71]. For instance, they may go to greater lengths to avoid being detected by surveillance systems, because they know they are at greater risk for being marked as false positives. These psychological harms and physical restrictions are injustices concentrated on Black people from the knowledge that a false positive is more likely if one is Black, rather than the effect from a false positive actually occurring.

The classification parity metric (see Table 1) addresses the *concentrating injustice* issue by equalizing false positive rates and false negative rates for different groups. But even then, equal justice may not be served: a second reason equal accuracy does not guarantee equal justice is that these algorithms occur against a backdrop of an unjust basic structure.[15] The same action can exacerbate pre-existing harms in an unjust basic structure for one group, but not another. This is called *compounding injustice* [50, 52].[16] Suppose that besides accuracy, the PSPRA algorithm's error rates are also the same for white defendants and Black defendants, all at 20%. However, a 20% false positive rate likely leads to more injustices for Black communities than it does for white communities. Why? False positives translate to pretrial detention, and I will show in §6.1 that detaining a Black defendant harms Black communities more than detaining a white defendant harms white communities.[17] The basic intuition is as follows. Because of an unjust basic structure, majority-Black neighborhoods have disproportionately high incarceration rates compared to majority-white neighborhoods [73, 88S]. When even more people in majority-Black neighborhoods are incarcerated, overall incarceration rates become so high that the entire community suffers from the consequences, which include: psychological burdens, difficulties building up social capital, broken community networks, and more [87]. This further reinforces an unjust basic structure. Essentially, existing injustices in the basic structure are compounded by false positives for Black communities, whereas this compounding effect does not apply to white communities (generally) because the basic structure is not unjust for them. Hence the same errors can impact different groups differently,

---

[15]See also [34, 54, 111] for discussions of how unjust conditions can lead to unjust outcomes even with error rate parity in the context of algorithmic fairness. Taking into account existing injustices is in keeping with non-ideal theory in philosophy.
[16]See [76, 80] for variations of the compounding injustice described here; sometimes also referred to as cumulative injustice.
[17]Black communities are also harmed by false negatives, especially those that translate into violent crimes. However, the false negative rate for both groups is identical in this scenario, suggesting that the additional harms caused by false negatives are unlikely to be disproportionate. I will further discuss the tradeoffs between pretrial release and public safety in §6.2.

because they may compound injustices for marginalized groups. Equal accuracies can once again serve unequal levels of justice.

What is the difference between *concentrating injustice* and *compounding injustice*? Both lead to increased injustices for a certain group. However, *concentrating injustice* introduces injustices that occur at higher rates (such as from an elevated false positive rate) for a certain group independent of any existing injustices. In contrast, *compounding injustice* introduces injustices that exist only because existing injustices for a certain group are exacerbated for that group. Figure 1 illustrates these distinctions; an equal accuracy rate for Black and white defendants is assumed for each scenario.

Notably, in all the scenarios, there are harms to individuals (i.e., from false positives). The first scenario simply has no group-level injustices because of equal error rates and a just basic structure. In contrast, the fourth scenario has an elevated false positive rate for Black defendants, *and* an unjust basic structure of society. This is the current situation with the COMPAS PSPRA algorithm. Here, equal accuracy rates do not guarantee equal justice due to both *concentrating injustices* and *compounding injustices*. First, a higher false positive rate for Black defendants [8] concentrates the injustices of false positives towards Black defendants. Second, the harms from false positives compound existing injustices for Black communities in a way that it does not for white communities. Hence, it is important to recognize that both *concentrating injustices* and *compounding injustices* are mechanisms by which equal accuracy does not entail equal justice. A solution that only takes into account *concentrating injustices* (i.e., classification parity [27] only fully addresses scenario two) fails to capture compounding effects from an unjust basic structure of society.

Overall, equalizing accuracy for groups has questionable value. Unequal justice may result because of *concentrating injustices* and *compounding injustices*. In fact, equalizing accuracy as the sole focus for fairness can even be detrimental. In the process, injustices that can only be addressed with unequal accuracy rates may be overlooked.

4.3 What Should We Value in Criminal Justice?

I have shown that measuring PSPRA algorithms' accuracy is problematic. Furthermore, even if we could perfectly measure accuracy, equal accuracy does not necessarily achieve the equal justice that a fairness notion should produce. Equal accuracy is flawed as a metric and value. Thus fairness notions that depend on equalizing some variation of accuracy, like the notions in Table 1, are undesirable.

However, is there room for fairness notions not based on accuracy in criminal justice settings? One might find it intuitive to discuss PSPRA algorithmic accuracy in the criminal justice setting because PSPRA algorithmic accuracy tries to balance individual liberty and public safety, two values clearly important in the criminal justice context.

But there are other values besides public safety and individual liberty that are important in criminal justice contexts. Police officers are (supposedly) not to enter homes without search warrants because we value privacy [81]. Spouses are (generally) legally protected from testifying against each other because we value the institution of marriage [12]. We have protections against double

jeopardy because we value protecting individuals from excessive state power [60].

Moreover, we forsake a certain degree of public safety in order to uphold these other values. We could likely catch more crimes if police could search homes without warrants. Similarly, we could likely punish more perpetrators of crime if we forced spouses to testify against each other, and tried individuals for the same crimes twice. Thus it is perfectly reasonable to balance public safety against other values during crime *detection* and crime *punishment*. Then it should be just as reasonable during crime *prediction* (as with PSPRA algorithms) to consider values beyond public safety. The next few sections describe such a value — relational equality — and what the most fair PSPRA algorithm would look like with fairness in terms of relational equality instead.

5 RELATIONAL EQUALITY

To the extent that equal accuracy is meant to capture the same procedure being applied to two groups, it can be thought of as a form of *procedural* justice.[18] Thus far, the criticism against equal accuracy has been on both *procedural* (as a metric) and *substantive* (as a value) grounds. Moving forward, I argue on *substantive* justice grounds that unlike equal accuracy, relational equality best embodies the point of equality, and is especially apt for considering racial fairness.

5.1 Overview of Relational Equality

There are two types of equality. The first treats people equally with respect to goods and/or opportunities [31, 185]. For instance, perhaps equality has been achieved if everyone has the same income or if all groups have the same accuracy score. We can refer to this type of equality as distributional equality. But as described earlier, there are drawbacks to conceptualizing equality as such: even if an algorithm is equally correct for everyone, that does not necessarily mean it produces just outcomes. In contrast to treating people equally, the second type of equality focuses on people relating as equals [103]. The distribution of material goods does not necessarily have to be equal for people to relate as equals. In fact, some people (such as those with disabilities) may require more of a certain resource in order to relate as an equal.

Relational equality belongs under the second type of equality.[19] Core to relational equality is democratic citizenship [11] and an ethos of mutual respect [106]. To be treated as equals, people need to be able to relate to each other as moral equals in a democratic society [5, 313]. Though this involves ensuring everyone has access to a basic level of certain (tangible and intangible) resources, including guaranteeing basic socio-economic necessities, the distribution of material goods is not the focus. Rather, power relations are the direct concern of relational equality. Relational equality "views equality as a social relationship" rather than a "pattern of distribution" [5, 312].

---

[18]For a discussion of procedural vs substantive justice, see [111].

[19]Besides relational equality, there are many other theories of equality belonging to egalitarianism, a school of thought in political philosophy based on equality [10]. Section 3 in [18] gives an overview of egalitarianism geared towards algorithmic fairness discussions.

	Perfectly Just Basic Structure in Society	Unjust Basic Structure in Society
False positive and negative rates are **equal** (and nonzero) for Black and white defendants	1. Equal Group Justice	3. Unequal Group Justice via Compounding Injustice
False positive rates are **higher** for Black defendants than for white defendants	2. Unequal Group Justice via Concentrating Injustice	4. Unequal Group Justice via Compounding Injustice + Concentrating Injustice

Figure 1: Risk scenarios created by *concentrating injustice* and *compounding injustice*

5.2 Elizabeth Anderson's Relational Equality: Three Types of Basic Capabilities

Philosopher Elizabeth Anderson's specific theory of relational equality translates the broad principle of relating as moral equals in a democratic society into distinctive requirements.[20] Anderson describes a set of negative and positive requirements that must be met for people to relate as equals in a democratic society. Evaluating the impact of PSPRA algorithms on these requirements for different racial groups is how we can use relational equality as a fairness notion.

According to Anderson, the negative requirements of relational equality are based on reciprocity; individuals must only act upon principles that are mutually justifiable [5, 313]. This translates to a lack of oppression in all its forms, including domination, exploitation, marginalization, and violence [5, 109]. Anderson also details a set of positive requirements for relational equality. This means building a democratic community rather than a hierarchical one [5, 313]. People must have access to basic capabilities that enable them to freely and fully "participate and enjoy the goods of society, and to participate in democratic self-government" as moral equals [5, 315].[21]

The basic capabilities in the positive requirements of relational equality are particular capabilities, rather than the set of all possible capabilities. For example, Anderson notes that the capability to be good at playing cards is not a basic capability that everyone needs access to, because one's status in civil society should be independent of one's card-playing prowess [5, 317]. Furthermore, while literacy is important for engaging with others in civil society as equals, understanding languages other than English, or arcane literary texts, is not needed in the United States [5, 319]. Hence a Ph.D. in English is not a basic capability that everyone needs, whereas a working proficiency in English is. Overall, relational equality does not call for equality of everything. Rather, relational equality requires access to specific basic capabilities that enable people to relate on equal terms with each other in society.

Anderson's basic capabilities also correspond to three important roles people take on: 1) as humans; 2) as participants in a system of social cooperation; and 3) as citizens [5, 317-8]. For example, to interact as equals, we require food and shelter (as humans); a basic education (as participants in social cooperation); and equal voting rights (as citizens).

5.3 Relational Equality and Algorithms

Algorithmic decision-making systems can alter access to the basic capabilities central to relational equality. Consider LA County's housing algorithm, which decides who of LA County's homeless population receives housing based on cost-benefit analysis [33, 75]. The problem is that LA County does not have enough housing for everyone, so the algorithm leaves a certain subset of the homeless population unhoused — those that do not pass the cost-benefit analysis [33, 75]. The individuals left unhoused by the algorithm are deprived of a basic capability: housing in some form. That makes it harder for them to relate as equals with others in society on basic human terms. Notably, giving everyone an equal probabilistic chance in receiving housing (ie, a random lottery)[22] does not solve the problem for relational equality; instead, housing scarcity must be fixed because the basic capability of housing must be available for everyone.

Finally, relational equality is especially relevant to racial fairness in PSPRA algorithms. As I will show in §6.1, incarceration via PSPRA algorithms affects not only defendants', but also in some cases their communities', access to basic capabilities. Moreover, relational equality is especially pertinent to racial fairness because relational equality is fundamentally concerned about which groups have power and which do not, and has no room for social hierarchies that oppress marginalized groups. Because of its focus on power relations, it is a framework expansive enough to encompass not only the socioeconomic and material harms, but also the status harms, defining racial injustices in the United States. Hence we should care deeply about relational equality as a value, especially when discussing race, because the United States' current racial injustices collectively constitute relational *in*equality.

6 AFFIRMATIVE PSPRA ALGORITHMS

What are the implications of using relational equality as the fairness notion in PSPRA algorithms, rather than existing statistical notions? In this section, I argue that the most racially fair version of a PSPRA algorithm – based on impact to relational equality – would be an Affirmative PSPRA algorithm. An Affirmative PSPRA

---

[20]Although I refer to Anderson's theory of relational equality simply as "relational equality" for clarity, Anderson calls it "democratic equality" [5, 313].
[21]This belongs to a broader framework termed the capabilities approach. See Martha Nussbaum [79] and Amartya Sen [92].

[22]This could be thought of as an equalized odds approach. See [47].

algorithm would have a more lenient risk threshold for Black defendants, similar in construction to race-based thresholds [28, 5].[23] Because Affirmative PSPRA algorithms treat otherwise facially similar Black and white defendants differently, one may argue that it is *procedurally* unjust.[24] However, I reiterate that this paper argues on the grounds of *substantive* justice, and specifically relational equality; treating like cases differently is not generally inconsistent with *substantive* justice, as Anderson argues in the context of affirmative action [6]. Moreover, this is not an argument for using PSPRA algorithms in general; in fact, perhaps abolishing pretrial detention may be the most fair [42, 71, 86]. Rather, I argue that where PSPRA algorithms are used, an Affirmative PSPRA algorithm would be the most racially fair version.

6.1 Relational Equality and PSPRA Algorithms

Incarceration decisions from PSPRA algorithms negatively affect access to the basic capabilities needed for relational equality, especially for Black communities. If pretrial detention is recommended, that can mean between 5 and 200 days of incarceration for a defendant [68].[25] Incarceration creates barriers to basic capabilities needed for relational equality under all three of Anderson's spheres of existence: as humans, as participants in a system of social cooperation, and as citizens. These harms are magnified for Black communities because of pre-existing higher rates of incarceration (a product of an unjust basic structure).

My analysis draws from disproportionately high existing rates of incarceration for majority-Black neighborhoods, compared to majority-white neighborhoods, owing to an unjust basic structure [73, 88S]. Many community members being incarcerated magnifies the harms from PSPRA algorithms. This is not to say individual defendants are not harmed regardless of race; rather, specific neighborhoods with high incarceration rates are harmed disproportionately. Moreover, though this paper focuses on Black communities because they have the highest overall incarceration rate [73, 84], the actual implementation will depend on city-specific incarceration rates, which may mean my analysis applies to other communities beyond Black communities. Finally, I emphasize that relational equality concerns individuals relating as equals, not communities. However, when an increased rate of incarceration has an effect on the community, it harms the ability of individuals in the community to access basic capabilities. This is why communities are relevant.

6.1.1 As Humans.
According to Anderson, it is impossible for people to relate as human beings of equal standing if some do not have adequate access to basic necessities like food, water, and health care [5, 317]. People need a basic level of physical and psychological safety before they can engage with the rest of society on equal terms as everyone else.

Suppose a PSPRA algorithm recommends pretrial detention for a defendant, and they are detained. Incarceration often means unjustified physical and psychological harm. Physical harms include physical assault from other inmates and jail staff [40], an overcrowded jail environment [38], and poor health due to low-nutrition food and disease outbreaks [22, 63, 91]. Pretrial detention also presents psychological harms to defendants in the form of a higher suicide rate [35, 3], poor availability of mental (and physical) health care [105], and post-traumatic stress disorders (including for formerly incarcerated individuals) [107], to name a few.

These psychological consequences impact not only defendants, but their communities, and especially Black communities. When a significant number of community members are incarcerated, as is the case for many majority-Black neighborhoods today, these communities can experience substantial psychological consequences. Children experience "serious psychological consequences" when separated from incarcerated parents [87, 1284]. Researchers even go as far to model incarceration as a "contagion" infecting neighborhoods, finding that individuals in neighborhoods with high incarceration rates are more likely to suffer from mental health disorders *independent* of whether they had actually been incarcerated themselves [49]. By incarcerating more Black community members, a PSPRA algorithm recommending pretrial detention for Black defendants would exacerbate access barriers to basic capabilities for Black communities.

6.1.2 As Participants in Social Cooperation.
Another set of Anderson's basic capabilities, including an adequate education and the ability to freely choose an occupation [5, 318], is necessary for people to relate as equal participants in a system of social cooperation. People should be able to contribute to society through their skills and/or talents — whether through the market or otherwise — and receive the benefits of such cooperation to make a living.

When a PSPRA algorithm incarcerates a defendant, that reduces defendants' job prospects through life disruptions and an increased chance of pleading guilty and future conviction. During detention, defendants face "varying levels of disruption" in "employment, financial situation, residential stability, and issues relating to dependent children" [57, 12]. By trial time, pretrial-incarcerated defendants are more likely to be found guilty than those released pretrial, even when correcting for other causal factors [29, 202], and plead guilty almost three times faster than non-detained defendants [83]. Likely explanations include accepting a plea bargain because of "the immediacy of being physically free" [67], being at rock bottom because of incarceration's harms [97], and generally being in a relatively weaker bargaining position, particularly for those "charged with less serious crimes and with no prior offenses" [97, 203]. Pleading guilty — almost always leading to a conviction — then translates to worse job opportunities down the road [1, 29].

When many Black community members are already incarcerated, further incarceration introduces and/or exacerbates community-level harms that make it difficult for individuals to relate as equals with others in the larger system of social cooperation. The immediate family of an incarcerated defendant suffers from social and financial strain, and when this is the case for *many* families in majority-Black neighborhoods, it is "harder for residents to form expansive networks that are most adept at producing social capital"

---

[23]"Affirmative algorithm" also appears in legal discussions [16, 24, 56]. I present the first comprehensive *ethical* defense of Affirmative PSPRA algorithms.

[24]However, whether it is generally possible for an algorithm to be completely procedurally just (treating like cases like) is disputed. See [4, 111].

[25]To simplify my analysis of how PSPRA algorithms affect relational equality, I assume that algorithm decisions are final. In reality, a judge makes the final decision from risk scores, which makes a difference (see [2, 96]) and necessitates jurisdiction-specific Affirmative PSPRA algorithms.

[87, 1282-3]. Social networks like churches and neighborhood associations are disrupted when there is a "mass movement of adults between the neighborhood and prison" [87, 1285]. This social capital is important for finding jobs, among other opportunities in society. Hence a PSPRA algorithm recommending pretrial detention for Black defendants would further deny basic capabilities for Black communities.

6.1.3   As Citizens. According to Anderson, to have equal relations as citizens, everyone must have more than just equal political rights. Being stigmatized or labeled as an outcast by other members of society makes for social hierarchies [5, 318]. People must be seen and heard on equal terms.

When PSPRA algorithms decide to detain defendants pretrial, that restricts defendants' ability to vote, and stigmatizes individuals. While those who are detained pretrial are legally eligible to vote, lack of access to voter registration forms, applications for absentee ballots and/or polling places, and information on deadlines creates pervasive "de facto disenfranchisement" for citizens who are incarcerated [88]. Studies have also found that pretrial detention — even without an ensuing criminal conviction — is enough to stigmatize an individual [102]. In fact, social stigma has been cited as one of the reasons to shift from pretrial detention to more concealable electronic monitoring (albeit with its own host of issues) [72, 144]. This is concerning for relational equality, because social hierarchies can be reinforced by the stigma associated with pretrial detention.

Partly because of its associated stigma, pretrial detention makes it even harder for Black individuals to relate with equal status to others. Incarceration still disproportionately affects Black communities today [77]. When even more Black defendants are held behind bars pretrial, it feeds into a public perception of associating Blackness with criminality, despite a lack of criminal conviction during pretrial detention. Similar to the impact of racialized mass incarceration (and likely fueled by the media), such a perception "perpetuates racial subordination, worsens blacks' disadvantage, and stigmatizes African Americans as criminal deviants" [94, 210]. Such stigma against Black people is a significant barrier to relational equality and further compounds existing injustices towards Black people. For example, Blackness (for urban youth) is often associated with criminality, which is a disadvantage when seeking employment, housing, and good schools [94, 210]. Stigma prevents Black people from being seen and heard on equal terms as everyone else in society, reinforcing a social hierarchy and relational inequality.

Overall, because of an unjust basic structure, PSPRA algorithms' detention recommendations negatively impact relational equality for Black communities more so than white communities. Perhaps a natural conclusion here is to reject pretrial detention completely (e.g. as [42, 71, 86] have called for), obviating the need for PSPRA algorithms. Would that not improve relational equality the most? That could very well be the case. But to reiterate the scope of this paper: I argue within the constraints of PSPRA algorithms being used — as they are across the United States today [23, 48] — for what the most fair version of PSPRA algorithms would be based on relational equality. That is what I address next.

6.2   Affirmative PSPRA Algorithms

The most racially fair PSPRA algorithm, based on relational equality, must be one that takes into account the disproportionate harm of pretrial detention on Black communities' access to basic capabilities. Namely, such a PSPRA algorithm must detain fewer Black defendants than the status quo. One way to do so is to lower the risk score (or accordingly adjust the risk threshold) for Black defendants; let us call such an algorithm an Affirmative PSPRA algorithm. In some ways this is similar to the proposal of race-specific risk thresholds [28, 5]. What distinguishes Affirmative PSPRA algorithms from these proposals is its basis in relational equality, as opposed to satisfying a statistical metric like classification parity, which requires race-specific risk thresholds to meet its objective of equal error rates [27, 13].

Concretely, consider a PSPRA algorithm like COMPAS or PSA. An Affirmative PSPRA algorithm takes the risk score output by that algorithm and for Black defendants, lowers the risk score slightly. It seeks to give a slightly more lenient risk score to a Black defendant, compared to a white defendant with the same number of prior arrests.[26] In doing so, an Affirmative PSPRA algorithm releases, instead of detains, substantially more Black defendants than would a non-Affirmative PSPRA algorithm.

By lowering risk scores for Black defendants, an Affirmative PSPRA algorithm necessarily reduces false positives at the expense of more false negatives for Black defendants. But this is actually reasonable if what we care about is fairness, and specifically, relational equality as fairness. Restricted individual liberty from false positives imposes barriers to relational equality that are especially pronounced for Black communities, compared to white communities. False positives undermine relational equality in a way that false negatives do not; false negatives, for example, do not impede access to job opportunities for specific communities the way being incarcerated does. Hence lowering false positives is a mandate of relational equality.

But why is introducing more false negatives justifiable, especially with the public safety risks they introduce?[27] I give two reasons. First, it is possible that the public safety costs of increasing false negatives are offset by the reduction of false positives. Both false positives and false negatives carry public safety risks, as discussed in §3.2. That means lowering false positives while increasing false negatives does not necessarily mean an increase in public safety risks overall. Some of the risks are introduced earlier (by false negatives) instead of later (by false positives).

Second, this paper's focus is on fairness. While I showed in §4.3 that there is room for values beyond public safety in the criminal justice system, the specific tradeoff between fairness and public safety is a much larger conversation that should likely be left up to jurisdiction-level democratic decision-making processes. Hence I only show what relational equality as fairness requires. As long as the public safety risks from false negatives do not undermine

---

[26]The actual numbers for the threshold will depend on what is viable and will likely vary from jurisdiction to jurisdiction. And of course, the number of arrests for an individual is not the same as the number of crimes they committed, and likely inflated for Black defendants.

[27]We might be especially worried if "the cost of releasing a high-risk defendant mostly fall on members of that defendant's community (e.g., when violent crime in a community is mostly committed by members of that community)" [27, 14].

relational equality, they are justifiable under relational equality as fairness. How might public safety risks undermine relational equality? Access to basic capabilities must be substantially impacted by recidivism instances. In particular, the number and type of recidivism instances must be enough to substantially change people's physical and perceived sense of public safety in the neighborhood — two human-level basic capabilities that would be most relevant. But nonviolent offenses such as shoplifting and drug usage account for a vast majority of pretrial recidivism instances. In Washington D.C., for example, only 1% of defendants released pretrial recidivated violently [36, 12]. While undesirable, introducing more cases of shoplifting is unlikely to alter a community's physical and psychological safety to a significant extent.[28] Hence the false negatives introduced are unlikely to affect access to the basic capabilities required for relational equality in a meaningful way. This is true regardless of where the recidivism instances occur, including majority-Black neighborhoods. Overall, insofar as fairness and relational quality as fairness is concerned, an increase in false negatives to reduce false positives for Black defendants should not be alarming.[29]

Even if we could predict the future, and knew which defendants would recidivate if they were released pretrial, an Affirmative PSPRA algorithm would still release some Black defendants who would commit a crime (for example, all Black defendants who would commit a minor crime).[30] Pretrial detention poses such a barrier to relational equality that doing otherwise would be unfair from a relational equality perspective.

Affirmative PSPRA algorithms also stand in contrast to the important and often-suggested goal of correcting for data bias (e.g. [61, 99]). §3.1 discussed how racial bias in policing practices likely inflates arrest rates in Black defendants' data [70, 78, 82], making it difficult to measure algorithmic accuracy. A natural solution would be to correct for such data bias. But even if there was perfectly unbiased arrest data, relational equality would still require Affirmative PSPRA algorithms to release more Black defendants; instead, data bias needs to be *overcorrected* for.

In fact, in contrast with traditional PSPRA algorithms, Affirmative PSPRA algorithms do not have maximizing accuracy as a goal for Black defendants. Exactly correcting for racial bias in police arrests would increase the accuracy of the algorithm the most. Specifically, consider a hypothetical algorithm that does not consider any variables that can be tied to racial bias in arrests, such as zip codes, arrest history, and even conviction history, but that manages to achieve high (> 70%) accuracy. An Affirmative PSPRA algorithm would still not be satisfied with this comparatively accurate algorithm as long as there is a significant false positive rate for Black defendants. Rather, an Affirmative PSPRA algorithm would be more lenient for Black defendants in order to reduce the false positives rate, even at the expense of increasing false negatives and lowering the overall accuracy for Black defendants.

The upshot is that equalizing accuracy is *neither necessary nor sufficient* for fairness based on relational equality. First, as §4.2 suggests, equal accuracy is *not sufficient* for relational equality as fairness. Even if PSPRA algorithmic accuracy is equal for Black and white defendants, the effects of pretrial detention in an unjust basic structure are disproportionately more harmful to Black communities' access to basic capabilities. Second, equal accuracy is also *not necessary* for relational equality as fairness. I discussed earlier that Affirmative PSPRA algorithms lower the overall accuracy for Black defendants. In contrast, the high accuracy for white defendants would be unchanged. Thus white and Black defendants would have unequal accuracy for relational equality as fairness via Affirmative PSPRA algorithms.

It is actually desirable that fairness based on relational equality – an Affirmative PSPRA algorithm – de-emphasizes accuracy. Constructing a future that is more fair and just must necessarily produce outcomes that are inaccurate according to today's unjust basic structure with relational *in*equality; otherwise, we are merely holding a mirror to current injustices and magnifying them.

7 CONCLUSION

In conclusion, current statistical notions of algorithmic fairness are undesirable because the foundation of equalizing accuracy does not capture the substantive point of equality in fairness in the first place. Relational equality is a fruitful alternative framework. When applied to the PSPRA algorithmic setting, relational equality requires Affirmative PSPRA algorithms, where risk scores are lowered for Black defendants. These Affirmative PSPRA algorithms show that equalizing accuracy is neither necessary nor sufficient for relational equality.

ACKNOWLEDGMENTS

---

[28]Moreover, offenses like shoplifting violate civic obligations that Black people arguably do not have the duty to uphold precisely because the system of social cooperation that demands civic obligations is unjust [94, 213-220]. In contrast are violations of moral obligations (i.e., assault) [94, 213].

[29]The caveat is unless false negatives are increased to such high levels, which I do not attempt to specify here, that access to basic capabilities is affected.

[30]This is, of course, with the same caveat as earlier: the introduced public safety risk must be capped to a level that does not affect relational equality.

REFERENCES

[1] Amanda Agan and Sonja Starr. 2017. The Effect of Criminal Records on Access to Employment. *American Economic Review* 107, 5 (May 2017), 560–564. https://doi.org/10.1257/aer.p20171003

[2] Alex Albright. 2019. If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions. *Working Paper* (2019), 39.

[3] Michelle Alexander. 2020. *The new Jim Crow: mass incarceration in the age of colorblindness, tenth anniversary edition* (10th anniversary edition ed.). New Press, New York. OCLC: 1138682569.

[4] Ali Alkhatib and Michael Bernstein. 2019. Street-Level Algorithms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3290605.3300760

[5] Elizabeth S. Anderson. 1999. What is the Point of Equality? *Ethics* 109, 2 (1999), 287–337. http://www.jstor.org/stable/2989479

[6] Elizabeth S Anderson. 2002. Integration, Affirmative Action, and Strict Scrutiny. *New York University Law Review* 77, 5 (Nov 2002), 1195–1271.

[7] Julia Angwin. 2016. Sample COMPAS Risk Assessment: COMPAS "Core". https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE

[8] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. How We Analyzed the COMPAS Recidivism Algorithm. https:

//www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=9WLmsuYtAV9xfiOMxs8XxIDtYNohEqSV

[9] Advancing Pretrial Policy & Research (APPR). 2021. About the Public Safety Assessment: How it Works. https://advancingpretrial.org/psa/factors/

[10] Richard Arneson. 2013. Egalitarianism. In *The Stanford Encyclopedia of Philosophy* (Summer 2013 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[11] Richard J. Arneson. 2010. Democratic Equality and Relating as Equals. *Canadian Journal of Philosophy Supplementary Volume* 36 (2010), 25–52. https://doi.org/10.1080/00455091.2010.10717653

[12] American Bar Association. 2021. Marital Privileges. https://www.americanbar.org/groups/litigation/publications/litigation_journal/2018-19/summer/marital-privileges/

[13] Michelle Bao, Angela Zhou, Samantha Zottola, Brian Brubach, Sarah Desmarais, Aaron Horowitz, Kristian Lum, and Suresh Venkatasubramanian. 2021. It's COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks. *CoRR* abs/2106.05498 (2021). arXiv:2106.05498 https://arxiv.org/abs/2106.05498

[14] Chelsea Barabas. 2020. Beyond Bias: Re-imagining the Terms of "Ethical AI" in Criminal Law. *Georgetown Journal of Law & Modern Critical Race Perspectives* 12, 2 (2020), 83–111. https://doi.org/10.1093/oxfordhb/9780190067397.013.47

[15] Ruha Benjamin. 2019. *Race after technology: abolitionist tools for the new Jim code.* Polity, Medford, MA.

[16] Jason R Bent. 2020. Is Algorithmic Affirmative Action Legal? *The Georgetown Law Journal* 108, 4 (April 2020), 51.

[17] 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* 50, 1 (July 2018), 3–44. https://doi.org/10.1177/0049124118782533

[18] Reuben Binns. 2018. Fairness in Machine Learning: Lessons from Political Philosophy. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 149–159. https://proceedings.mlr.press/v81/binns18a.html

[19] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2021. The Values Encoded in Machine Learning Research. *CoRR* abs/2106.15590 (2021). arXiv:2106.15590 https://arxiv.org/abs/2106.15590

[20] Renée Jorgensen Bolinger. 2021. Explaining the Justificatory Asymmetry between Statistical and Individualized Evidence. In *The Social Epistemology of Legal Trials*, Jon Robson and Zachary Hoskinks (Eds.). Routledge, 60–76.

[21] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of Machine Learning Research*. PMLR, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html ISSN: 2640-3498

[22] Eddie Burkhalter, Izzy Colón, Brendon Derr, Lazaro Gamio, Rebecca Griesbach, Ann Hinga Klein, Danya Issawi, K. b Mensah, Derek M. Norman, Savannah Redl, Chloe Reynolds, Emily Schwing, Libby Seline, Rachel Sherman, Maura Turcotte, and Timothy Williams. 2021. Incarcerated and Infected: How the Virus Tore Through the U.S. Prison System. *The New York Times* (April 2021). https://www.nytimes.com/interactive/2021/04/10/us/covid-prison-outbreak.html

[23] Electronic Privacy Information Center. 2021. *AI and Human Rights: Criminal Justice System.* Technical Report. https://epic.org/ai/criminal-justice/index.html

[24] Anupam Chander. 2017. The Racist Algorithm? *Michigan Law Review* 115, 6 (2017), 1023–1045. https://michiganlawreview.org/wp-content/uploads/2017/04/115MichLRev1023_Chander.pdf

[25] Alex Chohlas-Wood. 2020. *Understanding risk assessment instruments in criminal justice.* Technical Report. Brookings Institution. https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

[26] Alexandra Chouldechova and Aaron Roth. 2018. The Frontiers of Fairness in Machine Learning. *arXiv:1810.08810 [cs, stat]* (Oct. 2018). http://arxiv.org/abs/1810.08810 arXiv: 1810.08810.

[27] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv:1808.00023 [cs]* (Aug. 2018).

[28] 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* (Halifax, NS, Canada) *(KDD '17)*. Association for Computing Machinery, New York, NY, USA, 797–806. https://doi.org/10.1145/3097983.3098095

[29] Will Dobbie, Jacob Goldin, and Crystal S. Yang. 2018. The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review* 108, 2 (Feb. 2018), 201–240. https://doi.org/10.1257/aer.20161503

[30] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through Awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (Cambridge, Massachusetts) *(ITCS '12)*. Association for Computing Machinery, New York, NY, USA, 214–226. https://doi.org/10.1145/2090236.2090255

[31] Ronald Dworkin. 1981. What is Equality? Part 1: Equality of Welfare. *Philosophy & Public Affairs* 10, 3 (1981), 185–246. https://www.jstor.org/stable/2264894

[32] Equivant. 2019. Practitioner's Guide to COMPAS Core. http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf

[33] Virginia Eubanks. 2018. High-Tech Homelessness. *American Scientist* 106, 4 (2018), 230. https://doi.org/10.1511/2018.106.4.230

[34] Sina Fazelpour and Zachary C. Lipton. 2020. Algorithmic Fairness from a Non-ideal Perspective. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* ACM. https://doi.org/10.1145/3375627.3375828

[35] International Association for Suicide Prevention (IASP). 2007. *Preventing suicide in jails and prisons.* World Health Organization, Geneva, Switzerland. OCLC: 669968499.

[36] Pretrial Services Agency for the District of Columbia. 2020. Congressional Budget Justification and Performance Budget Request: Fiscal Year 2021. https://www.psa.gov/sites/default/files/FY2021%20PSA%20Congressional%20Budget%20Justification_0.pdf

[37] The Leadership Conference Education Fund. 2022. Pretrial Risk Assessments. https://civilrights.org/edfund/pretrial-risk-assessments/

[38] Gaby Galvin. 2017. Understaffed and Overcrowded: State Prisons Crippled by Budget Constraints, Bad Leadership. *U.S. News & World Report* (July 2017). https://www.usnews.com/news/best-states/articles/2017-07-26/understaffed-and-overcrowded-state-prisons-crippled-by-budget-constraints-bad-leadership

[39] P. Garg, J. Villasenor, and V. Foggo. 2020. Fairness Metrics: A Comparative Analysis. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE Computer Society, Los Alamitos, CA, USA, 3662–3666. https://doi.org/10.1109/BigData50022.2020.9378025

[40] Dave Gilson. 2016. What We Know About Violence in America's Prisons. *Mother Jones* (2016). https://www.motherjones.com/politics/2016/06/attacks-and-assaults-behind-bars-cca-private-prisons/ Section: Crime and Justice.

[41] Erving Goffman. 1986. *Stigma : notes on the management of spoiled identity.* Simon & Schuster Inc, New York, New York.

[42] Vaidya Gullapalli. 2020. The Path To Liberation Does Not Go Through Pretrial Punishment. https://theappeal.org/the-path-to-liberation-does-not-go-through-pretrial-punishment/

[43] Lelia Marie Hampton. 2021. Black Feminist Musings on Algorithmic Oppression. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/3442188.3445929

[44] Leif Hancox-Li and I. Elizabeth Kumar. 2021. Epistemic Values in Feature Importance Methods: Lessons from Feminist Epistemology. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 817–826. https://doi.org/10.1145/3442188.3445943

[45] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a Critical Race Methodology in Algorithmic Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 501–512. https://doi.org/10.1145/3351095.3372826

[46] Nikole Hannah-Jones. 2016. Choosing a School for My Daughter in a Segregated City. *The New York Times* (Jun 2016). https://www.nytimes.com/2016/06/12/magazine/choosing-a-school-for-my-daughter-in-a-segregated-city.html

[47] Moritz Hardt, Eric Price, and Nathan Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Barcelona, Spain) *(NIPS'16)*. Curran Associates Inc., Red Hook, NY, USA, 3323–3331.

[48] Heather Harris, Justin Goss, and Alexandria Gumbs. 2019. *Pretrial Risk Assessment in California.* Technical Report. Public Policy Institute of California. 25 pages. https://www.ppic.org/publication/pretrial-risk-assessment-in-california/

[49] 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 (Jan. 2015), 138–143. https://doi.org/10.2105/AJPH.2014.302184

[50] Deborah Hellman. 2018. *Chapter 5. Indirect Discrimination and the Duty to Avoid Compounding Injustice.* Hart Publishing, 105–122.

[51] Deborah Hellman. 2020. Measuring Algorithmic Fairness - Virginia Law Review. *Virginia Law Review* 106, 4 (June 2020), 811–866. https://www.virginialawreview.org/articles/measuring-algorithmic-fairness/

[52] Deborah Hellman. 2021. Big Data and Compounding Injustice. *Journal of Moral Philosophy* (2021). https://doi.org/abstract=3840175 Forthcoming.

[53] Matt Henry. 2019. Risk Assessment: Explained. https://theappeal.org/the-lab/explainers/risk-assessment-explained/

[54] Jonathan Herington. 2020. Measuring Fairness in an Unfair World. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* ACM. https://doi.org/10.1145/3375627.3375854

[55] Tiana Herring. 2020. *Releasing people pretrial doesn't harm public safety.* Technical Report. Prison Policy Initiative. https://www.prisonpolicy.org/blog/2020/

11/17/pretrial-releases/

[56] Daniel E. Ho and Alice Xiang. 2020. Affirmative Algorithms: The Legal Grounds for Fairness as Awareness. https://lawreviewblog.uchicago.edu/2020/10/30/aa-ho-xiang/

[57] Alexander M Holsinger. 2016. *The Effects of Pretrial Detention on Self-Reported Outcomes*. Research Brief. Crime and Justice Institute. 19 pages.

[58] Human Rights Watch. 2020. US: Californians Should Reject Proposition 25. https://www.hrw.org/news/2020/07/29/us-californians-should-reject-proposition-25

[59] Kim Steven Hunt and Billy Easley II. 2017. *The Effects of Aging on Recidivism Among Federal Offenders*. Technical Report. United States Sentencing Commission, Washington, DC. 54 pages. https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2017/20171207_Recidivism-Age.pdf

[60] Cornell Law School: Legal Information Institute. 2021. Double Jeopardy. https://www.law.cornell.edu/wex/double_jeopardy

[61] Heinrich Jiang and Ofir Nachum. 2019. Identifying and Correcting Label Bias in Machine Learning. *arXiv:1901.04966 [cs, stat]* (Jan. 2019). http://arxiv.org/abs/1901.04966 arXiv: 1901.04966.

[62] Jay Jordan and Sam Lewis. 2020. California voters: Don't be fooled by law enforcement and the bail industry in 2020. *The Sacramento Bee* (Nov. 2020). https://www.sacbee.com/opinion/article246809702.html

[63] Impact Justice. 2020. *Eating Behind Bars: Ending the Hidden Punishment of Food in Prison*. Executive Summary. https://impactjustice.org/wp-content/uploads/IJ-Eating-Behind-Bars-ExecutiveSummary.pdf

[64] Erin I. Kelly. 2017. The Historical Injustice Problem for Political Liberalism. *Ethics* 128, 1 (2017), 75–94. https://doi.org/10.1086/692974 _eprint: https://doi.org/10.1086/692974.

[65] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. Human Decisions and Machine Predictions*. *The Quarterly Journal of Economics* (Aug. 2017). https://doi.org/10.1093/qje/qjx032

[66] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2016. Inherent Trade-Offs in the Fair Determination of Risk Scores. *arXiv:1609.05807 [cs, stat]* (Nov. 2016). http://arxiv.org/abs/1609.05807

[67] Juleyka Lantigua-Williams. 2016. Why Poor, Low-Level Offenders Often Plead to Worse Crimes. *The Atlantic* (July 2016). https://www.theatlantic.com/politics/archive/2016/07/why-pretrial-jail-can-mean-pleading-to-worse-crimes/491975/ Section: Politics.

[68] Patrick Liu, Ryan Nunn, and Jay Shambaugh. 2018. *The Economics of Bail and Pretrial Detention*. Technical Report. The Hamilton Project. 23 pages.

[69] Christopher T Lowenkamp, Marie VanNostrand, and Alexander Holsinger. 2013. *The Hidden Costs of Pretrial Detention*. Technical Report. Arnold Foundation. 32 pages.

[70] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2016.00960.x.

[71] M4BL. 2020. End to pretrial detention and money bail. https://m4bl.org/policy-platforms/end-pretrial-and-money-bail/

[72] Eric Maes and Benjamin Mine. 2013. Some Reflections on the Possible Introduction of Electronic Monitoring as an Alternative to Pre-trial Detention in Belgium. *The Howard Journal of Criminal Justice* 52, 2 (March 2013), 144–162. https://doi.org/10.1111/hojo.12008

[73] Marc Mauer. 2011. Addressing Racial Disparities in Incarceration. *The Prison Journal* 91, 3_suppl (Sept. 2011), 87S–101S. https://doi.org/10.1177/0032885511415227

[74] Sandra G. Mayson. 2019. Bias In, Bias Out. *The Yale Law Journal* 128, 8 (June 2019), 2218–2300. https://www.yalelawjournal.org/article/bias-in-bias-out

[75] Tanvi Misra. 2018. The Rise of 'Digital Poorhouses'. https://www.bloomberg.com/news/articles/2018-02-06/when-criminalizing-the-poor-goes-high-tech

[76] Andreas Mogensen. 2017. Racial Profiling And Cumulative Injustice. *Philosophy and Phenomenological Research* 98, 2 (Oct. 2017), 452–477. https://doi.org/10.1111/phpr.12451

[77] Ashley Nellis. 2021. *The Color of Justice: Racial and Ethnic Disparity in State Prisons*. Technical Report. The Sentencing Project. https://www.sentencingproject.org/publications/color-of-justice-racial-and-ethnic-disparity-in-state-prisons/

[78] Rebecca Neusteter and Megan O'Toole. 2019. *Every Three Seconds: Unlocking Police Data on Arrests*. text/html. Vera Institute of Justice. https://www.vera.org/publications/arrest-trends-every-three-seconds-landing/arrest-trends-every-three-seconds/findings Archive Location: https://www.vera.org/ Publisher: Vera.

[79] Martha Nussbaum. 2003. Capabilities as Fundamental Entitlements: Sen and Social Justice. *Feminist Economics* 9, 2-3 (Jan. 2003), 33–59. https://doi.org/10.1080/1354570022000077926

[80] Tim O'Brien. 2021. Compounding injustice: The cascading effect of algorithmic bias in risk assessments. *Georgetown Journal of Law & Modern Critical Race Perspectives* 13, 1 (2021), 39–83. https://doi.org/10.2139/ssrn.3694818

[81] Library of Congress. 2021. The Need for a Warrant to Conduct a Search or Seizure: Overview | Constitution Annotated. https://constitution.congress.gov/browse/essay/amdt4_3_2_1/

[82] National Academies of Sciences Engineering and Medicine. 2018. *Proactive Policing: Effects on Crime and Communities*. The National Academies Press, Washington, DC. https://doi.org/10.17226/24928

[83] Nick Petersen. 2020. Do Detainees Plead Guilty Faster? A Survival Analysis of Pretrial Detention and the Timing of Guilty Pleas. *Criminal Justice Policy Review* 31, 7 (Aug. 2020), 1015–1035. https://doi.org/10.1177/0887403419838020

[84] The Sentencing Project. 2018. *Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System*. Technical Report. https://www.sentencingproject.org/publications/un-report-on-racial-disparities/

[85] John Rawls. 1977. The Basic Structure as Subject. *American Philosophical Quarterly* 14, 2 (1977), 159–165. https://www.jstor.org/stable/20009663 Publisher: [North American Philosophical Publications, University of Illinois Press].

[86] René Reyes. 2021. Abolition Constitutionalism and Non-Reformist Reform: The Case for Ending Pretrial Detention. *Connecticut Law Review* 53, 3 (Sep 2021), 667–679.

[87] Dorothy E. Roberts. 2004. The Social and Moral Cost of Mass Incarceration in African American Communities. *Stanford Law Review* 56, 5 (2004), 1271–1305. https://www.jstor.org/stable/40040178 Publisher: Stanford Law Review.

[88] Danielle Root and Lee Doyle. 2018. *Protecting the Voting Rights of Americans Detained While Awaiting Trial*. Technical Report. Center for American Progress. https://www.americanprogress.org/issues/democracy/reports/2018/08/23/455011/protecting-voting-rights-americans-detained-awaiting-trial/

[89] Richard Rothstein. 2017. *The color of law : a forgotten history of how our government segregated America*. Liveright Publishing Corporation, a division of W.W. Norton & Company, New York London.

[90] Jonathan Rothwell and Douglas S. Massey. 2009. The Effect of Density Zoning on Racial Segregation in U.S. Urban Areas. *Urban Affairs Review* 44, 6 (April 2009), 779–806. https://doi.org/10.1177/1078087409334163

[91] Robert Salonga. 2021. Coronavirus: Santa Clara County jail infections keep soaring as officials ponder another batch release. *The Mercury News* (Jan. 2021). https://www.mercurynews.com/2021/01/13/coronavirus-santa-clara-county-jail-infections-keep-soaring-as-officials-ponder-another-batch-release

[92] Amartya Sen. 1985. *Commodities and Capabilities*. North-Holland, Amsterdam. New Delhi: Oxford University Press, 1987; Italian translation: Giuffre Editore, 1988; Japanese translation: Iwanami, 1988..

[93] Eliza Shapiro. 2021. New York Schools Are Segregated. Will the Next Mayor Change That? *The New York Times* (Jan. 2021). https://www.nytimes.com/2021/01/29/nyregion/nyc-mayoral-race-school-segregation.html

[94] Tommie Shelby. 2016. *Dark Ghettos: Injustice, Dissent, and Reform*. Harvard University Press.

[95] Jennifer L. Skeem and Christopher T. Lowenkamp. 2016. Risk, Race, and Recidivism: Predictive Bias and Disparate Impact. *Criminology* 54, 4 (Nov. 2016), 680–712. https://doi.org/10.1111/1745-9125.12123

[96] Megan T. Stevenson. 2018. Assessing Risk Assessment in Action. *Minnesota Law Review* 103 (2018), 303. https://doi.org/10.2139/ssrn.3016088

[97] Megan T Stevenson. 2018. Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes. *The Journal of Law, Economics, and Organization* (Sept. 2018). https://doi.org/10.1093/jleo/ewy019

[98] Cass R. Sunstein. 2019. Algorithms, Correcting Biases. *Social Research: An International Quarterly* 86, 2 (2019), 499–511. https://muse.jhu.edu/article/732187 Publisher: Johns Hopkins University Press.

[99] Tatiana Tommasi, Novi Patricia, and Tinne Tuytelaars. 2015. A Deeper Look at Dataset Bias. (2015), 19. https://arxiv.org/ftp/arxiv/papers/1505/1505.01257.pdf

[100] Sahil Verma and Julia Rubin. 2018. Fairness Definitions Explained. In *Proceedings of the International Workshop on Software Fairness* (Gothenburg, Sweden) *(FairWare '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3194770.3194776

[101] Jodi L. Viljoen, Melissa R. Jonnson, Dana M. Cochrane, Lee M. Vargen, and Gina M. Vincent. 2019. Impact of risk assessment instruments on rates of pretrial detention, postconviction placements, and release: A systematic review and meta-analysis. *Law and Human Behavior* 43, 5 (Oct. 2019), 397–420. https://doi.org/10.1037/lhb0000344

[102] Sarah Cusworth Walker and Jerald R. Herting. 2020. The Impact of Pretrial Juvenile Detention on 12-Month Recidivism: A Matched Comparison Study. *Crime & Delinquency* 66, 13-14 (June 2020), 1865–1887. https://doi.org/10.1177/0011128720926115

[103] Michael Walzer. 1983. Spheres of Justice: A Defence of Pluralism and Equality. *Philosophy* 59, 229 (1983), 413–415.

[104] Leif Wenar. 2021. John Rawls. In *The Stanford Encyclopedia of Philosophy* (summer 2021 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/rawls/

[105] Andrew P. Wilper, Steffie Woolhandler, J. Wesley Boyd, Karen E. Lasser, Danny McCormick, David H. Bor, and David U. Himmelstein. 2009. The Health and Health Care of US Prisoners: Results of a Nationwide Survey. *American Journal of Public Health* 99, 4 (April 2009), 666–672. https://doi.org/10.2105/AJPH.2008.144279

[106] Jonathan Wolff. 1998. Fairness, Respect, and the Egalitarian Ethos. *Philosophy & Public Affairs* 27, 2 (April 1998), 97–122. https://doi.org/10.1111/j.1088-4963.1998.tb00063.x

[107] Nancy Wolff, Gregory Mchugo, Jing Shi, Jessica Huening, and B. Christopher Frueh. 2015. Screening for PTSD among Incarcerated Men: A Comparative Analysis of Computer-Administered and Orally Administered Modalities. *Criminal Justice and Behavior* 42, 2 (Feb. 2015), 219–236. https://doi.org/10.1177/0093854814551601

[108] Alice Xiang. 2021. Reconciling Legal and Technical Approaches to Algorithmic Bias. *Tennessee Law Review* 88, 3 (Jan 2021), 649–724. https://ssrn.com/abstract=3650635

[109] Iris Marion Young. 1990. *Justice and the Politics of Difference.* Princeton University Press.

[110] Iris Marion Young. 2006. Taking the Basic Structure Seriously. *Perspectives on Politics* 4, 1 (2006), 91–97. https://doi.org/10.1017/S1537592706060099

[111] Annette Zimmermann and Chad Lee-Stronach. 2021. Proceed with Caution. *Canadian Journal of Philosophy* (2021), 1–20. https://doi.org/10.1017/can.2021.17