A Dilemma for Criminal Justice Under Social Injustice

by

Deniz Ariturk

Program in Bioethics and Science Policy
Duke University

Date: November 13, 2019

Approved:

_____
Gopal Sreenivasan, Advisor

_____
Misha Angrist

_____
Walter Sinnott-Armstrong

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Arts in Bioethics and Science Policy in the
Graduate School of Duke University

2019

ABSTRACT

A Dilemma for Criminal Justice Under Social Injustice

by

Deniz Ariturk

Program in Bioethics and Science Policy
Duke University

Date: November 13, 2019

Approved:

_____
Gopal Sreenivasan, Advisor

_____
Misha Angrist

_____
Walter Sinnott-Armstrong

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of
Master of Arts in Bioethics and Science Policy in the
Graduate School of Duke University

2019

Abstract

A moral dilemma confronts criminal justice in unjust states. If the state punishes marginalized citizens whose crimes are connected to conditions of systemic injustice that the state has failed to alleviate, it perpetuates a further injustice to those citizens. If the state does not punish, it perpetuates an injustice to victims of crime whose protection is the duty of the criminal justice system. Thus, no reaction to crime by the unjust state appears to avoid perpetuating further injustice. Tommie Shelby proposes a new solution to this old dilemma, suggesting that certain theoretical and practical qualifications can save the unjust state from perpetuating injustice. He argues that punishment can be just even as society remains unjust, if it is: (a) administered through a fair criminal justice apparatus; (b) only directed at *mala in se* crimes; and (c) not expressive of moral judgment. In the first part of this thesis, I explore Shelby's solution to argue that while certain aspects of his theory are superior to alternative ones, it nonetheless fails to resolve the dilemma. In Part 2, I use a novel technological reform that promises to make criminal justice fairer, the AI risk assessment, as a case study to show why even punishment that meets Shelby's criteria will continue to perpetuate injustice as long as it operates under systemic social injustice. Punishment can only be just if society is.

Contents

Acknowledgements

I would like to thank my advisor, Gopal Sreenivasan, for his guidance, patience, and immensely helpful comments through many drafts of this thesis.

Introduction

On November 18, 1975, David Bazelon, then Chief Judge to the United States Court of Appeals for the District of Columbia Circuit, delivered the first of the J. Edgar Hoover Foundation Lectures. Bazelon dedicated his lecture to the "morality of the criminal law," which he believed was compromised by severe societal injustices that drive disadvantaged citizens to crime. He was deeply disturbed by the fact that most criminal defendants come from poor and socially marginalized communities, and that criminal law was largely unmoved by this reality. Bazelon argued that punishing people when societal injustice is the root of their criminal behavior is immoral, and called upon the members of the legal and political community—namely, all of us, to recognize and move to address this issue:

> "In my opinion, it is simply unjust to place people in dehumanizing social conditions, to do nothing about those conditions, and then to command to those who suffer, "Behave—or else!" The overwhelming majority of violent street crime, which worries us so deeply, is committed by people at the bottom of the socioeconomic-cultural ladder—the ignorant, the ill-educated, and the unemployed and often unemployable. I cannot believe this is coincidental. Rather, I must conclude that those people turn to crime for reasons such as economic survival, a sense of excitement or accomplishment, and an outlet for frustration, desperation, and rage. We cannot produce a class of desperate and angry citizens by closing off, for many years, all means of economic advancement and personal fulfillment for a sizeable part of the population, and thereafter expect a crime-free society." (Bazelon 1975: 401-402)

Bazelon's words, while powerful, have had no real legal impact, even as social injustice continues to pervade society over forty years later. As they were when Bazelon gave his lecture, American prisons today are disproportionately comprised of the most disadvantaged citizens in society. A 2014 study found that the pre-incarceration income of incarcerated citizens concentrates at the lowest end of the national income distribution, and that the median income of incarcerated people is 41 percent lower than that of their nonincarcerated co-citizens (Kelly 2018: 150), thus seeming to confirm Bazelon's claim that "poverty appears to be a necessary, though

not a sufficient, condition for the occurrence of most violent crime" (Bazelon 1975: 403). Prison populations are also disproportionately made up of historically marginalized communities. Two thirds of the people in prison today are people of color. Black non-Hispanic adult males are almost seven times more likely to be incarcerated than non-Hispanic white men, and Hispanic men are three times more likely to be incarcerated than white men. Over a third of Black men without high school diplomas are currently in prison (Kelly 2018: 151). In the face of these and many similarly disturbing statistics, the link between injustice and crime appears undeniable, and the morality of our current punishment practices dubious.

This thesis engages with the undeniable link between social and criminal justice, and the dubious morality of punishment under social injustice. I address the same problem that troubled Bazelon during his time as Chief Judge: the injustice of commanding, through punishment, that the victims of severe social injustice "behave," rather than addressing those social injustices. Bazelon's words capture a persisting moral dilemma for criminal justice in unjust societies like the US. In what is to come, I argue that the criminal justice practices of a society, particularly its state-imposed punishment, cannot be justified as long as severe injustice affects the lawful life prospects of its citizens. I outline this moral dilemma before responding to Shelby (2016), who presents one of the most recent and what I take to be the strongest responses to it. I raise three problems with Shelby's framework to show that it fails to fully address the injustice of punishment under conditions of social injustice. I then use a recent criminal justice reform, algorithmic risk prediction, to show why Shelby's response to the dilemma is insufficient, and why social justice is a necessary step in the road to justified punishment. Reforms that narrowly target criminal justice practices and leave social injustice unaddressed cannot make punishment just, no matter how technologically advanced, efficient, fair, or accurate they may be.

My argument that social justice is a necessary condition for just punishment has significant practical implications for how we shape not only our criminal justice practices, but also our social institutions more broadly. If we want punishment to be morally justified to all citizens, we need to adopt a preventative justice approach that addresses the societal causes of crime rather than a punishment-oriented scheme that targets the victims of those causes.

Part 1: An Old Dilemma, A New Solution, Three Problems

"I believe that there can be no truly just criminal law in the absence of social justice—in other words, you can't have one without the other."
—David Bazelon, former Chief Judge of the United States Court of Appeals for the District of Columbia Circuit

An Old Dilemma

Criminal justice under conditions of systemic social injustice raises a dilemma for the state: If the state punishes citizens whose crimes are connected to conditions of injustice the state has failed to alleviate, then it commits a further injustice to those citizens. If, on the other hand, the state refuses to punish those citizens to avoid further marginalizing them, it perpetuates an injustice to the victims of crime whose protection is the very duty of the criminal justice system. The state thus perpetuates injustice in both punishing and refraining from punishing crime. Is there a way out of this "moral predicament," as Watson (2015) calls it, apart from the alleviation of the background injustice? Or, is a just criminal law unattainable in the absence of social justice, as Judge Bazelon (1975) argued over 40 years ago?

Some contemporary legal scholars side with Bazelon, arguing that there is no way out of the dilemma for the unjust state. It cannot avoid committing further injustices. Victor Tadros (2009) posits that by creating criminogenic circumstances through distributive injustice, the state becomes complicit in the crimes of the poor and is responsible for their destructive effects. As such, the state cannot justifiably hold the poor responsible when the poor commit crimes, because doing so would mean that the state acts "as a judge when it ought to be a co-defendant" (410). Antony and Robert Duff (2001) similarly argue that the present practice of criminal justice in the US cannot be justified to disadvantaged defendants who are excluded from the political community. By hypocritically committing the same wrongs it criminalizes such as deceit, theft, and unjustified violence, the state loses its moral standing to condemn crime.

Since punishment involves condemnation, these scholars argue, state punishment under conditions of injustice is never just. In this way, they echo Bazelon's sentiment that the source of injustice in the criminal law is the "conflict between certain moral pretenses and practices" of criminal punishment (1975: 389). However, this does not mean that that state should never punish. Tadros and Duff argue that in cases of violent crime that threaten the safety of the citizenry, the state may retain the right to punish. Thus, state punishment under conditions of systemic social injustice is sometimes justifiable on crime control grounds. Still, this justification does not negate the injustice committed against those who are punished. Rather, when the injustice caused by allowing crime to go unpunished exceeds that caused by punishing the disadvantaged, the state may permissibly commit the lesser of two evils. As Duff puts it, the state in this situation has a "highly qualified justification of criminal punishment as a necessary but…morally tainted enterprise" (2001: 200). Here lies the moral predicament: the state must either perpetuate this morally tainted albeit necessary enterprise, or allow harm to vulnerable (and often disadvantaged) citizens. Neither option is desirable or just.

A New Solution

Tommie Shelby (2016) proposes a way out of the dilemma. He agrees with Tadros and Duff that by perpetuating conditions of intolerable injustice, the state loses its moral standing to condemn the crimes committed by the unjustly disadvantaged. When a state fails to keep injustice above a tolerable minimum, it loses its legitimate authority to create civic duties for its citizens through law. A state without legitimate authority also lacks the moral standing to condemn law-breakers: it cannot condemn citizens for breaking laws they have no obligation to obey. Crucially, however, Shelby does not believe that this loss of standing renders all criminal punishment unjust. Criminal punishment need not express moral condemnation, and can be justified under a

consequentialist, harm-prevention rationale. I will explain the two parts of his argument in turn before going on to inspect if it holds up under scrutiny.

The unjust state cannot condemn the crimes of the unjustly disadvantaged, Shelby argues, because it lacks "right-to-be-obeyed legitimacy," or "legitimate authority," for short (229)[1]. Legitimate authority is the state's "right to create obligations for others" through law (230). This right allows the state to criminalize actions that are not wrongs in themselves and demand that citizens not commit them. Many illegal acts fall under this category, such as laws that prohibit speeding, jaywalking, shoplifting, drug possession, tax evasion, welfare fraud, and participating in the underground economy. These acts do not constitute moral wrongs in themselves, but are deemed unacceptable by the laws of a particular state. As such, their criminal status depends on the legitimate authority of that state. Such laws do not rest on any *a priori* moral duty, but rather on the relationship of authority between the state and its citizens. Citizens are required to obey these laws "because of the *source* of the rules, not because of their *substance*" (230). Importantly, only certain sources, namely, those with legitimate authority, can justifiably demand obedience to the laws they issue.

In liberal democracies like the US, Shelby argues, legitimate authority depends on a fundamental requirement of justice: the principle of reciprocity between citizens and the state. According to this principle, citizens owe loyalty to the state's laws and institutions in return for the basic goods and services that social cooperation makes possible (20). When the principle of reciprocity is satisfied, citizens have a "civic duty" to comply with the state's laws. Shelby adopts Rawls' standard of "constitutional essentials" to determine, at minimum, what goods and services reciprocity necessitates (214). The constitutional essentials include "freedom of speech, conscience, assembly, and association; the right to vote and run for office; the right to due process

[1] All bare page citations are to Shelby (2016).

and judicial fairness…freedom of movement, free choice of occupation, and formal justice," and finally, "a social minimum that secures the basic material needs of citizens." These essentials "establish the political legitimacy of a social order by publicly affirming the equal status of all citizens under the rule of law." When the state fails to meet a minimum standard of justice by providing all citizens with these constitutional essentials, it fails to satisfy the principle of reciprocity and loses its right to subject citizens to its laws and impose civic duties.

In American society, where "the disadvantaged are regularly subjected to such stigmatizing and demeaning forms of servitude as the ghetto poor have been," Shelby argues, "a bright line has been crossed" regarding injustice (215). Persistent conditions in the dark ghetto make this clear: Institutional racism exists in many social institutions, including employment and housing, and stigmatizes ghetto residents. Welfare benefits and social entitlements for the poor and unemployed are insufficient and at times wholly unavailable. Full-time employment often does not provide a living wage, even though the material conditions of the US economy could afford to ensure that it does. Public schools in many poor neighborhoods are inadequate to prepare children to flourish and act as a barrier to social mobility. Disproportionate policing and incarceration rates in ghettos create a constant cycling of people from ghetto to prison, and spread a "criminal-minded ethos" in those neighborhoods. Given all this, those who argue that even the ghetto poor must obey the law because they perceive US society to be "imperfect but reasonably just" (215) either rely on too lenient a definition of justice, or disregard the realities of the ghetto. Shelby asserts, and I agree, that the economic and social order in the US "relegates a segment of its citizenry to humiliating forms of exploitation," and as such "cannot reasonably expect allegiance from that oppressed group."

Having lost its legitimate authority and consequently its right to create obligations through law, the unjust state "has no moral basis for condemning disobedience to its laws as such,

particularly the disobedience of those unjustly disadvantage in society" (244). According to Shelby, this is simply because, "if condemnation of disobedience is to be apt, then the state must have legitimate authority" (239)[2].

By Shelby's account, then, like Tadros and Duff's, unjust states like the US have no moral standing to condemn lawbreakers. However, Shelby argues, lacking the moral authority to condemn crime need not create a moral dilemma for all state-imposed punishment. A state without legitimate authority can nonetheless retain the right to penalize certain crimes. This is because punishing crime (through threatening, penalizing, and neutralizing offenders) is distinct from condemning crime. Diverging from Tadros and Duff, Shelby rejects "penal expressivism," the theory that punishment necessarily expresses a moral judgment of the criminal actor (239).

According to penal expressivism, by punishing, we express public condemnation of criminal offenders. Shelby argues that this assumption is dubious; it is unclear why we would assume punishment serves a moral function, as opposed to simply "containing dangerous individuals or providing potential lawbreakers with an incentive to refrain from violating the law" (241). Of course, the debate around the purpose punishment *should* have, moral retribution being one of them, is difficult to settle. However, regardless of what punishment *should* accomplish, it is possible to conceive of a society that, like Shelby posits, simply acts as a "fair and a reasonably good deterrent or crime-control device" without any symbolic moral judgment. As such, moral condemnation and enforcement of penalties are not inseparable functions of punishment. The "moral pretenses" of criminal law, to use Judge Bazelon's words, need not include moral condemnation.

[2] Shelby adds that losing legitimate authority is not the only way the state's moral standing can be compromised. He accepts the complicity and hypocrisy claims raised by Tadros and Duff, respectively, as other reasons (244).

According to Shelby, distinguishing between condemnation of lawbreaking and penalties for lawbreaking explains why a seemingly contradictory pair of claims are in fact compatible: Serious structural injustice compromises "the state's authority to punish criminal offenders and its moral standing to condemn crimes," yet the same state may also "permissibly punish at least some legal violations, even some crimes perpetrated by the oppressed" (228). This latter right to penalize certain crimes fulfills the state's duty to protect the citizenry from harm, and is justified by a form of legitimacy that is distinct from legitimate authority. Shelby calls this form of legitimacy "enforcement legitimacy" (232).

Enforcement legitimacy refers to the state's right to impose penalties for crimes. As outlined above, the enforcement legitimacy of a state without legitimate authority is reduced in scope; it no longer covers crimes that are created by the state. In cases where the unjust state also lacks fair, effective, and humane criminal justice institutions, its enforcement legitimacy may be entirely void. However, if the criminal justice institutions that deliver the penalties are sufficiently fair,[3] the state can penalize certain crimes (244). Specifically, enforcement legitimacy in such cases can apply to crimes that are "serious moral wrongs" in themselves (such as murder or rape), also called *mala in se* crimes.

This right to penalize *mala in se* crimes gains footing from the state's obligation to defend the vulnerable from harm. Shelby considers this justification an extension of our moral right to "repress actions that seriously threaten our lives, freedom of movement, bodily integrity, or material being" (233) to a formal system of punishment. Understood as such, state-enforced punishment for such actions is not only permissible, but can also further the cause of justice by acting as a means to "redistribute burdens so as to reestablish equity" within the cooperative

[3] Of course, a state the is unjust often does not have fair criminal justice institutions. I will address this point later on.

scheme in society (235). Punishment in an illegitimate state functions solely to prevent crime, not to express a moral judgment. Distinguishing as such between the state's moral standing to condemn crime and its right to penalize certain crimes, Shelby argues, resolves the dilemma at stake. The state can simultaneously protect the vulnerable from harm and punish criminal offenders without perpetrating any additional injustice, as long as it punishes without condemning, and restricts itself to crimes that are serious moral wrongs.

It may appear that Shelby's framework nonetheless involves an injustice to the unjustly disadvantaged who are punished for committing "serious moral wrongs": how is it that their punishment is justifiable, when the punishment of other law-breakers is not? According to Shelby, this is because *mala in se* crimes are part of a normative order that is separate from and prior to the civic and legal order. These moral rules, or natural duties, bind individuals as moral agents, not citizens. Thus, the hold of natural duties, unlike that of civic duties, is not determined by the legitimacy of the state people live in (219). All moral agents, be they citizens of a legitimate or illegitimate state, are bound by their natural duties to one another. For the unjustly disadvantaged, then, the permissibility of criminal acts depends not on their legality but rather their prior moral value. Even a deeply unjust state could punish these crimes (but not morally condemn them, having lost its moral standing to condemn by failing to establish a just background structure), in order to avoid serious, irreversible harm to victims. Doing so would not perpetuate further injustice to the unjustly disadvantaged, as they owe loyalty to their natural duties even if not to the unjust state. The dilemma appears resolved.

On its face, Shelby's account seems to address all forms of injustice involved in the original dilemma: No injustice is done to victims, because crimes that cause harm to are punished. No injustice is done to offenders who break laws created by the unjust state, because those are no longer punishable offenses. Finally, no injustice is done to offenders who are

punished for *mala in se* crimes, because in committing those crimes, offenders have failed their natural duties as moral agents.

On closer scrutiny, however, this framework leaves room for significant injustices. In the next section, I will raise three problems for Shelby that suggest that his account does not truly resolve the dilemma. First, he does not account for crimes that are not moral wrongs that could nonetheless cause serious harm. Second, his consequentialist harm-prevention rationale for punishment neglects the harms punishment inflicts on marginalized offenders. And third, Shelby's fairness requirement, which his argument needs to solve the dilemma, is not met in most societies, and would not resolve the dilemma even if it were. I will address each problem in turn to show that the dilemma of criminal justice without social justice cannot be fully resolved by Shelby's qualifications.

Three Problems for Shelby

1. Crimes That Are Not Serious Moral Wrongs Can Cause Serious Harm

My first concern with Shelby's framework is that it appears to equate infliction of serious harm with commission of serious moral wrongs, when the two are in fact intersecting yet distinct categories of acts. On Shelby's account, punishment serves to prevent acts that cause serious harm to others. However, punishable offenses are not defined as "acts that cause serious harm," but rather as serious moral wrongs, or *mala in se* crimes. If it turns out that certain crimes that are not serious moral wrongs nonetheless cause serious harm, Shelby's account would not tell us how the state should respond to them.

Criminal acts that violate our natural duties constitute serious moral wrongs. They are thus covered by Shelby's solution to the dilemma, as they fall under the enforcement legitimacy of the unjust state. Shelby lists the duties to not be cruel, to not cause unnecessary suffering, and to show mutual respect among the most fundamental natural duties (219). It is unclear, however,

if the collection of serious moral wrongs exhausts all possible acts that could cause harm. Can the vulnerable only be harmed by acts that are serious moral wrongs?

I argue that the answer to this question is no. Citizens who do not commit moral wrongs may nonetheless seriously harm others. I will focus here on one way this can happen, namely through the collective effects of individual crimes. Sometimes, an individual crime may remain below the threshold of "serious moral wrong" and thus be unpunishable by the state, while many instances of the same crime could eventually harm other citizens. This would create a problem for Shelby, as harm would occur and punishment would be unjustified. In exploring this misalignment, I will refer to two crimes that Shelby argues could be justifiably committed by the unjustly disadvantaged: petty theft and gang membership.

Under Shelby's account, petty theft is a crime that cannot be punished by a state without legitimate authority. Shelby argues that certain financial crimes, including petty theft, are permissible under conditions of injustice, as long as they do not cause serious harm to others. For instance, robbing a Walmart at gunpoint would constitute a punishable offense, as it would violate the natural duty to not cause unnecessary suffering. In this instance, the state could justifiably punish the criminal by relying on its enforcement legitimacy, which gives the state the right to "repress actions that seriously threaten our lives, freedom of movement, bodily integrity, or material being" (233). Merely shoplifting from the same Walmart, however, would be a permissible crime. It causes no threat to the lives, freedom of movement, or bodily integrity of other citizens. Because it would cause a negligible loss to the corporation's profit, it would not constitute a threat to the "material being" of others, either. As such, it would not be covered by the enforcement legitimacy of the unjust state.

In fact, such minor financial crimes may not only be permissible, but at times morally praiseworthy in light of what Shelby calls the duty of justice. The duty of justice is a "corollary of

the value of justice itself," and refers to the citizenry and the state's collective *natural* duty to protect just institutions when they are in place, and to change them when they are not (57). For Shelby, "the very idea of social justice presupposes the duty of justice," which suggests that complying with certain laws when they perpetuate injustice might itself be unjust. This idea is lies at the heart of acts of civil disobedience, and is likely what James Baldwin (2011) had in mind when he wrote that "Some laws should not be obeyed," or Martin Luther King, Jr. (1963) did when he spoke about our "moral responsibility to disobey unjust laws."

Shelby points out that the duty of justice may justify financial crimes when the economic order in society is deeply unjust. In many societies including the US, the legal economy severely marginalizes and exploits certain communities like the ghetto poor, and does not allow their members to achieve decent living standards through full-time work. This leads to a "self-reproducing exploitative relationship" (197) between poor and affluent citizens, wherein the unjust power differential between the two groups allows the affluent to maintain or increase their power advantage through the sacrifices the poor must continually make for the affluent. This relationship keeps the poor unable to exit poverty and traps them in the "roles of maids, nannies, dishwashers, maintenance workers, and so on" that serve the affluent without offering any means for skills enhancement or promotion to the poor. For marginalized citizens who are faced with a deeply unjust and inescapable economic order, the refusal to participate in the legal economy may not merely be the only viable option to make a decent living, but also an act of civil disobedience. It may express the dissenters' refusal to be exploited by the unjust economic system that, in its current form, resembles "state-sponsored labor camps or workhouses for the black poor" (199). Acts like shoplifting or tax evasion may offer the ghetto poor the means to make a living, but also push the social order toward justice by creating a disturbance for powerful, advantaged citizens

that may compel them to recognize the injustice the system is perpetuating and take action to address it.

Under Shelby's account, then, certain economic crimes are not only permissible for the unjustly disadvantaged, but also a requirement of their natural duty of justice. This suggests that these crimes do not fall under the category of offenses the unjust state may punish. However, if enough people were to commit such financial crimes, the economy of the entire country would suffer. This would in turn harm the material being not only of large corporations and the state, but of all citizens who depend on the stability of the economic order, including the unjustly disadvantaged[4]. Shelby points to this possibility as well, but does not discuss how it can be prevented. After all, there are presumably enough people who currently live under conditions of severe injustice and thus have the right to commit financial crimes that they could together seriously disrupt the entire economy. Additionally, acts of dissent tend to facilitate a recognition of injustice and a change towards justice when they create disruptive effects through their collective force. When there aren't enough dissenters to create disruption, dissent is likely to go unnoticed. Economic disruption may thus be not only a permissible but a desirable consequence of financial crimes from the justice perspective. From the harm perspective, however, the same acts could cause serious harm to the citizenry, including the unjustly advantaged.

A similar point can be made about other crimes Shelby considers not to be covered by the unjust state's enforcement legitimacy, like gang membership. Shelby argues that, "in light of the hazards of participating in gang culture, recruiting children into gangs shows insufficient concern for the weak and vulnerable" (220). But, he goes on, "given the advantages of concerted group action, participating in gangs may be a defensible and effective means to secure needed income."

[4] The consequences of the financial crises of past decades, including the 2008 stock market crash, should serve as an indicator of just how harmful serious economic disruptions tend to be for marginalized citizens who do not have a safety net.

Gang membership, when social injustice makes alternative means of securing a livelihood unavailable, is a justified choice that cannot be punished by the illegitimate state.

However, while individual participation in gangs is justifiably left to the judgment of consenting adults, increases in gang membership overall risk harming the vulnerable. Given the "hazards of participating in gang culture," it is plausible that, if most people in disadvantaged neighborhoods took advantage of their right to participate in gangs, members of those neighborhoods could be physically or mentally threatened and even seriously harmed. No single individual's choice to participate in a gang would be responsible for this effect (unless their membership involved the commission of a violent crime), but the collective effect of many individual choices would indeed create an unsafe environment where others fear for their "lives, freedom of movement, bodily integrity, or material being."

In Shelby's account, then, the vulnerable face risk of harm that would go unpunished. The choice of unjustly disadvantaged citizens to commit minor financial crimes or join gangs, although permissible and perhaps morally righteous individually, could collectively harm the citizenry, including the unjustly disadvantaged. In other words, acts that are serious moral wrongs (i.e., punishable offenses) and those that cause serious harm do not align perfectly. Unless we are willing to argue either that only harms caused by *mala in se* crimes fall under the state's purview (in which case the vulnerable would be left unprotected), or that all acts that cause harm constitute a serious moral wrong (in which case the disadvantaged would be penalized for many crimes that injustice engenders and justice requires), the misalignment between harms and moral wrongs seems irreconcilable. In this area of misalignment, the state will either have to punish those it has no right to punish, or fail to protect marginalized citizens it has treated unjustly. The dilemma persists.

Before moving on, let me address a potential response to my objection that may save the unjust state from necessarily perpetuating injustice in this area of misalignment. In raising my objection, I have assumed that protecting the vulnerable from harm requires punishment of criminal offenders. Perhaps this is a misrepresentation of Shelby's solution of the dilemma. Certain harms, like those caused by financial crimes, could presumably be alleviated by the state without relying on the punishment of criminals. In fact, punishment in such cases is unlikely to redress the financial harm caused to victims. Instead, the state could avoid perpetuating injustice by providing monetary compensation to the victims who bear the burden of these crimes. This may sound like an economically unfeasible solution. However, given the exorbitant costs of running the criminal justice apparatus—especially of arresting, sentencing, and imprisoning people for non-violent crimes—it may turn out be an economically advantageous alternative to current practices[5]. In fact, having to bear the burden of the financial crimes of the unjustly disadvantaged may incentivize the state to correct the exploitative structure of the economic system, and free the ghetto poor from the economic restrictions that drive them to these crimes in the first place.

It is less clear if a similar argument could be made for gang membership, where the harms to the victims are not financial in nature. Physical and psychological harms cannot be undone through financial restitution. However, the state may nonetheless have resources at its disposal to offer victims, such as medical care or psychological counseling. While such services cannot take back the harms done to victims, they may nonetheless be a more effective restitution method than the punishment of offenders. An even better alternative would be the institution of

[5] In 2017, it cost the federal government an average $36,300 to incarcerate a prisoner for one year, or $95 for one day (https://www.federalregister.gov/documents/2018/04/30/2018-09062/annual-determination-of-average-cost-of-incarceration). "Petty theft," on the other hand, is defined as the theft of property valued under $400-$500 (https://www.legalmatch.com/law-library/article/petty-theft-law.html).

preventative interventions in marginalized communities that correct for the injustices that incentivize gang membership.

Shelby's account can be vindicated, then, if: (a) punishment of offenders and protection of citizens are not equated; and (b) the state has at its disposal alternatives to punishment that can address or prevent the harms done to victims. In fact, applying these two qualifications to Shelby's framework have implications for the necessity (and justifiability) of punishment broadly, including in cases of *mala in se* crimes. Separating the link between punishment and prevention of harm should make us rethink how we ought to respond not only to nonviolent offenders, but also to violent ones. I will return to this idea in the next two sections.

2. Harm Prevention Rationale of Punishment Neglects Harm to Marginalized Offenders

Above, I gave two examples of lawbreaking that would not be punishable under Shelby's regime but could nonetheless cause harm to vulnerable victims. In this section, I will focus on the unjust harms marginalized offenders are left vulnerable to under his framework. I will argue that, by focusing only on victims of crime, Shelby's harm-prevention justification of punishment ignores the harm done to marginalized citizens who get punished for *mala in se* crimes. In many instances of punishment, this harm could be prevented if the state chose to address the severe background injustice violent offenders are exposed to, and as such renders punishment unjust.

Shelby's category of punishable crimes is narrow. Under his restricted account of the unjust state's enforcement legitimacy, many disadvantaged offenders who serve prison sentences under current laws would no longer be punished. These lawbreakers would thus be protected from the injustice perpetuated by punishment under severe social injustice. Offenders who violate their natural duties and cause serious harm to others, on the other hand, would be subject to punishments including incarceration, even if they are unjustly disadvantaged. Shelby argues that

this differential treatment of offenders is justified through a utilitarian, harm-prevention rationale. Incarceration would presumably incapacitate and rehabilitate violent offenders who commit *mala in se* crimes and deter potential ones[6], thereby reducing harm to the vulnerable (even though, as we saw, serious harm is not caused exclusively by *mala in se* crimes). However, a significant shortcoming of this utilitarian justification is that it accounts solely for harm to the victims of crime (i.e., the "vulnerable"), not its perpetrators[7]. I argue that we should include perpetrators of violent crime[8], especially those who were victimized by severe social injustice, in our harm considerations. If we acknowledge the unjust harm that even Shelby's narrow conception of punishment inflicts on offenders, we will see that his framework cannot adequately protect unjustly disadvantaged citizens from harm. This in turn might push us to more diligently look for alternatives to punishment in our harm prevention efforts.

By separating condemnation from punishment in his response to the dilemma, Shelby appears to designate condemnation as the true source of injustice in punishment. However, in the absence of a just background structure, punishment that does not express moral condemnation can nonetheless cause unjust harm to disadvantaged citizens who are punished. Such punishment perpetuates injustice even when, as Shelby suggests, it is delivered through a fair criminal justice apparatus and only directed at *mala in se* crimes. This is because the harms of punishment are

[6] I leave aside the question of whether incarceration actually achieves these ends, even though I believe that, in its current state, it often does not. The astonishingly high recidivism rates of released prisoners, for instance, suggest that US prisons are criminogenic rather than rehabilitative environments that are not effective at crime prevention.

[7] This drawback is not unique to Shelby's framework; other utilitarian theories of punishment that rely on a harm prevention rationale overlook the harms done to prisoners by framing their goal as preventing harm to victims. If their goal was re-framed as preventing harm to all citizens, punishment would appear as a less justified solution in many instances of lawbreaking.

[8] It should also be noted that "victims" and "perpetrators" of violent crimes are sometimes the same individuals, coming from the same marginalized communities. Recognizing that today's victim may become tomorrow's perpetrator shows that in discourses of crime and punishment, there aren't neat divisions between innocent and guilty persons when those people are collectively victimized by the unjust background conditions of society.

vast, and many of them persist even after Shelby's fairness criteria are satisfied (I address shortcomings of these criteria in the next section). There is, above all, a fundamental harm confining people in cages inflicts on them, no matter how humane the cages may be.[9] Beyond its inherent dehumanizing nature, incarceration also harms the physical, psychological, and social wellbeing of prisoners. It separates citizens from their loved ones, communities, homes, and jobs; confines them to extended periods of solitude and monotony; and exposes them to the harsh public stigma attached to criminals[10]. In other words, some of the very harms punishment is meant to prevent, such as harms to our freedom of movement and material being, are inflicted on prisoners through punishment. These harms cannot be justified to individuals whose crimes were engendered by the failures of an unjust state to provide them with opportunities to lead a crime-free life.

I suspect that many will be opposed to this broader conception of harm. They will argue that there is a fundamental difference between the harm facing victims and perpetrators of harm: Victims are innocent and perpetrators are not. We should not be concerned with reducing harm to

[9] Take, for example, Norwegian prisons, which are internationally lauded for their "radical humaneness" (Benko 2015), short sentence lengths, and small inmate populations. While these prisons most likely meet Shelby's fairness criteria, Norwegian prison officials nonetheless recognize that "the stigma and the suffering lives [there], too" (Dreisinger 2016: 295). A Norwegian criminal justice professor recounts the time when, upon being asked if they would live in the nicest prison in the country rent-free for life after their sentence, prisoners responded with a resounding "never," reminding us that these institutions, no matter how humane they become, will always remain cages.

[10] Shelby recognizes that punishment may further stigmatize the unjustly disadvantaged, even when it does not involve condemnation. In the American context, it might perpetuate stereotypes of black criminality and violence, stereotypes for which the state's past and present criminal justice practices are in large part responsible for. He stresses that to prevent such stigmatization, "the state should make it clear that it penalizes, perhaps reluctantly, only to prevent unjust and harmful aggression, recognizing that it may be partly at fault for these wrongs" (249). I worry that such a public statement by the state, although symbolically significant, would do little to change the public perception of criminal offenders. Current discourses in the media and among the public suggest that people see criminal offenders as severely morally tainted. These discourses and perceptions are a great obstacle for a state that wishes to convey that its punishment does not express a moral judgment.

criminals, because they deserve the harm punishment inflicts. In fact, infliction of harm may be one of the very goals of incarceration for those who hold a retributivist notion of punishment. I do not intend to argue against retributivism here; the debate about the goal of punishment is not one I can settle. Instead, I will simply adopt Shelby's account of the acceptable function of punishment in a state without legitimate authority. As we saw, Shelby believes (and I agree) that a state without legitimate authority has lost the moral standing to morally condemn offenders. As such, punishment is justifiable only insofar as it serves to prevent serious harm to others; it is not supposed to depend on what the offender may or may not deserve. Even if *mala in se* crimes merit moral condemnation and the criminals who are punished for them deserve harm, the state does not have the moral standing to inflict it, and state-sanctioned punishment is not supposed to deliver it.

This is not to suggest that the unjust state should never punish violent offenders, or that all punishment of *mala in se* crime is unjust. In no way do I aim to disregard the harm victims of violent crime suffer. When punishment is the only or most effective means available of preventing such harm, I grant that punishment would be necessary. I seriously doubt, however, that this is always or even often the case. As we saw in the discussion of petty theft and gang membership in the previous section, punishment and protection from harm are not synonymous. The state often has tools beyond punishment at its disposal to address the harms caused by crime and even to prevent them. This applies to violent crime as well, although perhaps with a narrower scope. These tools could be used for providing restitution for past crimes and preventing future ones. Of course, the state can never address the full extent of harms caused to victims and their families by violent crime. Restitution in the form of financial or psychological assistance will fall significantly short of doing so. However, the same could be said for the potential of punishment to address past harm: even if the perpetrator were convicted to life in prison, his harms would not

be undone. A better alternative to trying to undo the harms of violent crime would thus be focusing on their prevention. There is good reason to believe that the state could significantly lower rates of violent crime by addressing the unjust background structure of society. Injustice plays a criminogenic role by reducing people's ability to avoid violent crime, which suggests that alleviating injustice would play a crucial role in crime prevention. These are the cases of violent crime I want to problematize: I argue that pursuing punishment is unjust when a less harmful alternative like the alleviation of injustice exists.

Shelby acknowledges that addressing injustice would reduce crime. He describes at length the criminal ethos that develops in ghetto neighborhoods, and the violence gang culture breeds. Citizens born into these neighborhoods are faced with incredible adversity that may severely limit their ability to avoid violent crime. As Shelby points out, "inequality, poverty, ghetto conditions, and low educational attainment are strongly correlated with (if not causes of) violent crime" (251). Given this, "criminal acts among the unjustly disadvantaged could be controlled through the establishment and maintenance of a more just basic structure." This would avoid injustice both to would-be criminals and would-be victims.

In fact, seen in the context of the duty of self-respect (221), classically understood "serious moral wrongs" are revealed to sometimes be a reasonable reaction to the harm and injustice faced by marginalized communities. According to Shelby, the duty of self-respect is required by the duty of justice in conditions where "freedom or perhaps even relief seems unattainable." This duty is an "ethic of resistance aimed at living with self-respect despite insurmountable injustice." In such dire situations, victims of injustice can affirm their equal moral worth as persons by resisting injustice. For example, they can fulfill their duty of self-respect by "standing up for oneself when one has been treated unjustly, rather than meekly acquiescing." I agree fully with Shelby on this point. For many people who face insurmountable injustice in

many parts of the world today, the duty of self-respect is one of the few available means of asserting agency, and must thus be championed.

However, asserting self-respect may involve violent responses to injustice. Christopher Lewis points out that "*malum in se* crime may be the most attractive (and sometimes the only) path to self-sufficiency and social standing in communities isolated from the mainstream economy." (2016: 153). For example, he argues, "in neighborhoods where legitimate ways to secure social status are scarce, violent crime, and the reputation it often comes with, can directly boost one's social standing," and "in low-income neighborhoods where the likelihood and costs of encountering a violent person are high, mutual respect can become a zero-sum game, leaving people with incentives to adopt a threatening demeanor and to behave in unfriendly, uncivil, and disrespectful ways towards others—often breaking the law in doing so." In such scenarios, people may cause serious harm to others in an effort to stand up for themselves in the midst of severe injustice, and consequently be subjected to the unjust state's punishment.

In an interview on the National Public Radio show *Fresh Air*, Sean Moore, a former dogfighter from Chicago, talks about the difficulty of avoiding immoral acts in marginalized neighborhoods (Kelly 2018: 89). Moore explains that he got involved in dogfighting because of the "bully guys" in his neighborhood. Becoming a successful dogfighter allowed Moore to escape what he calls the "negativity" he was born into. Dogfighting provided him with social status, which "means a lot" where he grew up, "especially on the negative side, because you don't want to walk down the street and be bullied, get your money took and beaten on as a punk in the neighborhood amongst a lot of criminal activity." Engaging in the morally problematic act of dogfighting allowed Moore to "walk through the neighborhood" without fear. In the end, Moore quit dogfighting because a young man named Julian got shot and killed by "gangbangers" for refusing to fight his dog. Moore's decision to quit what he realized was a violent, immoral

environment is admirable and was undoubtedly difficult. It also reveals just how unreasonable it is to expect people to avoid committing criminal acts when resisting to do so would mean risking their life. Moore could have ended up like Julian had he refused to comply with the "bullies" who demanded that he fight his dog.

The story of Sean Moore and Julian points to the role of injustice in many criminal actions of people who are victimized by severe injustice. Living in marginalized neighborhoods with low prospects for legally attaining social status and high incidences of violence makes it difficult for citizens to avoid committing serious moral wrongs. They may even commit these wrongs as a reaction to the injustice they have suffered. While this cannot diminish the harm their acts cause to victims, it should make us worried about considering the punishment of such disadvantaged lawbreakers justified—at least as justified as it would be if the perpetrators were not victims of injustice. Suggesting that Moore and others in similar circumstances deserve punishment for an act for which the alternative is risk of death is morally problematic. When the state's unjust structure leaves certain citizens in an environment of "negativity" where immoral action appears unavoidable on the road to self-respect and even safety, allowing the state to punish those citizens seems misguided, even if the action is an immoral one.

To be clear, I am not suggesting that solving the problem of social injustice would solve the problem of violent crime. Not all offenders of violent crime come from marginalized backgrounds, and not all unjustly marginalized citizens commit violent crimes. Attributing the criminal behavior of marginalized offenders entirely to their social circumstances would be a disrespectful dismissal of their moral agency. Injustice does, however, clearly play a role in limiting lawful alternatives for marginalized citizens and affects the makeup of the prison population. The prison population (including the population of violent criminals) in the US is, and historically has been, composed disproportionately of people who come from racially and

economically marginalized backgrounds.[11] While injustice is not the sole contributor to violent crime, it plays a large enough role to taint the entire practice of punishment. As long as criminal justice is used as a way to control broader social problems engendered by injustice the state has failed to address, punishment will continue to inflict unjust harms to the disadvantaged.

In the end, then, Shelby's approval of punishment, like Tadros and Duff's, is an argument for the lesser of two evils. It accepts the harm marginalized prisoners suffer as the casualty of the harm-prevention mission of punishment. However, such an argument cannot be just in circumstances where there exists a third alternative that is morally superior to the two evils in question. In our case, this alternative would be addressing the injustices that severely limit marginalized offenders' ability to avoid crime. I believe that this is a viable option for a country as wealthy and powerful as the US. The state's failure to pursue it is a political choice, not a necessity. As such, even if punishing violent offenders of marginalized backgrounds may sometimes be an effective method of crime control, it is not the least harmful one available to the state. The harms punishment inflicts on marginalized offenders is preventable and thus unjustifiable. The dilemma persists.

3. Shelby's Fairness Requirement is Not Met (and Would Not Resolve the Dilemma Even if it Were)

The final problem that remains unresolved in Shelby's account is the problem of unjust states with unfair criminal justice institutions. Shelby rightfully argues that an unjust state must, at a minimum, have a criminal justice apparatus that operates fairly if the state is to retain its

[11] This is not to say that prisoners make up the entirety of people who have committed violent crimes. There is a good chance that a disproportionate amount of racially and economically privileged violent offenders never get punished, either because they are never caught or because they can afford the best defense lawyers available. Alternatively, they may be likely to get shorter sentences than marginalized citizens for similar crimes. This, of course, further supports the argument that injustice plays a significant role in who ends up punished.

enforcement legitimacy. Specifically, a sufficiently fair criminal justice system would need to: (a) issue public warnings; (b) offer defendants adequate opportunity to defend themselves; (c) be impartial and evenhanded in applying the rules; and (d) issue penalties that are "humane and no more severe than necessary to deter the crime in question" (233). If this fairness requirement is not met, the punishment of *all* crimes, including those that are serious moral wrongs, is rendered illegitimate.

Shelby recognizes that, in the real world, many unjust states including the US do not meet the fairness requirement, nor do they seem likely to meet it in the foreseeable future. Such states "routinely mistreat innocent persons, criminal suspects, those with outstanding warrants, defendants, and convicts…through unjustified searches, racial profiling, police brutality, arbitrary and uneven enforcement, wrongful convictions, unfair sentences, and inhumane prison conditions" (228). In the US, black men disproportionately bear the burden of these unfair practices, resulting in a system of "racialized mass incarceration" (209). There is evidence that black people are "subjected to racial profiling and unjustified searches, exposed to gratuitous police violence and harassment, face racially biased juries, receive overly severe sentences, are subject to the arbitrary and excessive power of prosecutors, are not provided adequate legal counsel, and are not allowed to fully reintegrate into the political community after their sentences are served" (247).

What then, should be said of the state's right to punish in these societies? It appears that Shelby's reasonable yet unmet fairness requirement brings us back to the original dilemma: either the state has no right to punish until it establishes a fair criminal justice apparatus (in which case the vulnerable are left unprotected from harm), or law-breakers are punished unjustly. I suspect that Shelby does not attempt to solve the dilemma for such states, because he recognizes that there is no way for them to avoid injustice in their response to crime.

This is no small wrinkle. Shelby's response to the dilemma works only insofar as the criminal justice apparatus meets the fairness requirement. For him, "much will depend on whether the criminal justice system operates in a reasonably impartial and fair way, not on whether the state has the standing to condemn crime" (246). This is a natural extension of his position that punishment in unjust societies functions only to control crime, not to condemn. In other words, "holding someone accountable," such as through state-imposed punishment, depends "not on having the standing to condemn the wrong, but on having the standing to be an impartial judge of whether the accused committed the prohibited act." The importance of the "impartial judge" for Shelby is why, I believe, his most urgent reform suggestions target the criminal justice apparatus.

Many necessary criminal justice reforms follow directly from Shelby's fairness requirement. Currently, the American criminal justice apparatus falls significantly short of all of the criteria listed above. As such, many current US criminal justice practices would need to be abandoned and many new ones implemented for punishment to be justified. Some of the practices to be abolished include unjustified searches and racial profiling, judicial and prosecutorial bias, disproportionately long sentences that are not "justified by the need to deter the type of unjust conduct in question," and inhumane practices like extended periods of solitary confinement. Indigent defendants would need to be provided access to adequate legal counsel, and prison conditions would need be improved to meet standards of basic decency. The US currently denies ex-offenders many public benefits of citizenship, including "income subsidies, housing assistance, grants and loans for education, and unemployment insurance" (250). Felons in most US states are denied the vote during and after their time in prison. Such practices of discrimination and disenfranchisement would need to be abolished, for they not only fail the fairness requirement for being unnecessarily severe, but also perpetuate the unjust

marginalization of offenders in their communities. In fact, such penalties are likely to perpetuate criminal conduct, as disenfranchising ex-offenders and limiting their access to public benefits will further alienate them from society and make it more difficult than before for them to find lawful alternatives for survival. There is a retributivist tinge to such severe policies that go contrary to crime reduction motives; they appear not only to target crime control, but also to morally condemn and dehumanize offenders. And as we have seen, the unjust state has no moral standing to condemn in this way.

The criminal justice reforms proposed by Shelby are not insignificant. While they would require a significant overhaul of the system and are unlikely to gain political support in the near future, they are nonetheless more narrow, and thus presumably more realistic, goals than addressing the injustice that pervades society as a whole. These reforms, if realized, would certainly make punishment more morally palatable than its current implementation in the US. Crucially, however, they would fail to account for the impacts of the broader historical and institutional context of injustice as it relates to crime. As such, I argue that Shelby's fairness requirement is a beneficial metric for criminal justice-oriented reforms that nonetheless cannot render punishment just if systemic injustice persists in other parts of society.

Shelby himself points out that the reforms outlined above would be insufficient to address the full extent of injustice caused by the criminal justice apparatus. Doing so would require attempting to rectify the long history of marginalization caused by past unfair criminal justice practices. Shelby maintains that, if the unjust state is to maintain its enforcement legitimacy, "it should be making a good-faith effort to warrant and acquire the trust of the oppressed" (249), which decades of marginalization are likely to have eroded. Addressing the problem of criminal justice through the principle of rectification (12) would require the institution of policies that aim to "reconcile with, and make amends to, those it has wronged" in order to

"regain legitimacy in their eyes and acknowledge its role in creating the conditions under which the disadvantaged are tempted to turn to crime," (250). Among such policies would be educational and voluntary rehabilitative programs at prisons, and support programs for former prisoners during their reintegration into society.

However, while admirable and urgently needed, policies informed by the principle of rectification are nonetheless unlikely to prevent injustice to unjustly disadvantaged offenders under Shelby's account, for two reasons: First, they are institutionally broad interventions that are not related directly to the imposition of penalties. As such, they are not covered by Shelby's four fairness criteria, which narrowly target the administration of criminal penalties through the criminal justice apparatus. Rectification-oriented reforms like rehabilitative programs require recognition of the broader, interconnected injustices in society, such as lack of access to sufficient educational and financial opportunities to allow one to pursue a law-abiding life. While Shelby himself advocates for these restitutive reforms, his fairness criteria cannot secure that they get taken up by an unjust state that has enforcement legitimacy. The unjust state could justifiably impose penalties through the establishment of a minimally fair criminal justice apparatus, and neglect to implement rectification policies along the way. This is a concerning conclusion, as these policies appear crucial to my and Shelby's understanding of what justice requires.

This brings me to my second concern about rectification-oriented reforms to criminal justice: They are only directed at individuals who have already committed crimes. Those who benefit from them must simultaneously bear the burden of punishment. As mentioned in the previous section, these individuals have to go through the harms of being imprisoned, which living in tolerable levels of injustice and thus with sufficient opportunities to avoid crime could have saved them from. Because these rectification policies are necessarily backward-looking, they cannot serve a preventative function (except, perhaps, by lowering incidences of recidivism

by supporting offenders' re-entry into society). Because they are implemented through the criminal justice apparatus, they cannot address the criminogenic nature of injustice itself.

This is my greatest worry about the adequacy of the fairness requirement to justify punishment: Although Shelby's fairness requirement is currently unmet and the reforms necessary to meet it require substantial political will and time, these obstacles could nonetheless be overcome. I am not so sure about the promise of just punishment when injustice persists outside of the criminal justice apparatus. I disagree with Shelby's argument that punishment by the unjust state would not perpetuate injustice, as long as the fairness requirement is met. In Part 2, I will explore the debate around a recent criminal justice reform offering to make criminal justice fairer to show why. While this new reform could indeed be a "fairer" alternative to current practices, it falls short of addressing the injustice of punishment under social injustice. This is because the narrow, targeted reform is necessarily oblivious to the unjust background structure of society, which must be addressed for punishment to have legitimacy.

Part 2: Predicting Risk Under Injustice

The Risk Assessment Revolution

In Part 1, I argued that as long as the criminal justice apparatus operates under conditions of severe, systemic social injustice, even fairly administered punishment that is restricted to *mala in se* crimes will continue to constitute an injustice towards offenders who were victims of that injustice. This argument has practical implications for how we assess proposed criminal justice reforms that purport to make punishment more just by providing improvements to the fairness, efficiency, and accuracy of the criminal justice system.

One of the latest reforms offering this promise is the algorithmic risk assessment tool. This tool is considered the next generation of actuarial risk assessments that have a long history in the US criminal justice system, but had been out of fashion since 1970 (Garrett and Monahan 2018). Historically, these tools have used statistical models to assess an offender's likelihood of committing future crimes using variables that strongly correlate with recidivism in sample data. Included among these variables are socio-demographic information about the defendant such as past criminal history, education level, employment history, and neighborhood and family characteristics.

After falling out of fashion in the 1970s, risk prediction tools have recently returned, the algorithmic tool being the most technologically advanced. Risk assessment is now used across all stages of the criminal justice system and across all US jurisdictions, culminating in what academics have called a "risk assessment revolution" (Garrett and Monahan 2018: 1). Judges rely on risk scores in making decisions about pretrial detention, bail, sentencing, and parole. A risk score thus helps decide whether a defendant goes to prison, how long he goes to prison for, and whether he can be released. Proponents of risk prediction champion these tools as empirically backed methods to reduce recidivism and prison populations simultaneously through their ability

to differentiate between offenders who are safe enough to re-enter society and those who pose a significant enough threat to warrant incarceration. With the incorporation of artificial intelligence (AI), the tools can now presumably work on larger data sets more efficiently to provide more "accurate" predictions.

Not everyone is as enthusiastic about the AI risk assessment revolution, however. Over the past few years, the tools have come under intense scrutiny on fairness grounds. Most famously, a 2016 ProPublica article (Angwin et al.) reported race disparities in the outcomes of COMPAS, an AI risk assessment tool used widely in bail decisions. The researchers found that the algorithm was twice as likely to categorize a black defendant as high risk when he was in fact low risk, and more likely to categorize a white offender as low risk when he was in fact high risk. This disproportionate distribution of false positive and false negative errors along racial lines led to higher bail determinations for African American defendants than for white defendants charged with a similar crime. Academic articles have also pointed to similar race and gender bias in AI tools used in the criminal justice system, such as during predictive policing and facial recognition (Buolamwini and Gebru 2018; Lum and Johndrow 2016). These findings appear to confirm former US Attorney General Eric Holder's worry back in 2014 that risk assessment would "undermine our efforts to ensure individualized and equal justice" by exacerbating "unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society" (Angwin et al. 2016).

In response to these criticisms, researchers in academia and industry have devoted their efforts to eliminating bias from the tools through regulation and development of "fair, accountable, and transparent algorithms (FAT algorithms)" (Barabas 2019: 19). The ideal FAT algorithm would have the same predictive accuracy regardless of the defendant's characteristics. In doing so, the tools would presumably make the criminal justice apparatus fairer by providing a

better alternative to what the Model Penal Code calls the "notoriously imperfect" judgment of "responsible actors in every sentencing system," like judges and parole board members (Starr 2014: 815). FAT algorithms would protect us from both the implicit (or explicit) racial biases and the more minute cognitive failings[12] of criminal justice officials. They would serve as the "impartial judge" (246) that Shelby argues would make punishment of *mala in se* crimes justified under social injustice[13].

In what is to come, I challenge this last claim. Even if it were able to thoroughly satisfy Shelby's fairness requirement, algorithmic risk prediction would not justify punishment in an unjust society like the US. To understand the true costs and benefits of AI risk assessment tools, we must take apart the discourses of risk, crime, and safety in a society where most prisoners come from marginalized communities. We must move "beyond bias" (Barabas 2019) understood narrowly, and examine how risk assessment tools interact with the broader social framework they work in. Doing so will reveal the inadequacy of even the algorithmically perfect "impartial judge" to justify punishment until the systemic injustice that pervades other social institutions is alleviated.

[12] For instance, see Danziger et al. (2011) for how judges' hunger levels affect the number of favorable rulings they make.

[13] To be clear, I do not claim that Shelby's framework would require him to support AI risk assessment. In fact, I suspect that Shelby would be quite critical of this tool for reasons beyond fairness concerns. Among these would be the fact that the tools deem as high-risk many non-violent offenders, and that they have been used to support pretrial detention practices, particularly of marginalized communities. My focus here will not be on such failings of the tool. I aim instead to challenge the fairness rhetoric around risk assessment to show that addressing fairness issues alone cannot lead to a just punishment scheme. I hope to suggest by analogy that Shelby's account of crime and punishment, specifically his fairness criterion, suffer from the same shortcomings as the promise of fair risk assessment tools.

Three Justice Problems for AI Risk Prediction

Assessing AI risk prediction tools in the context of the historical and contemporary injustices of US society raises three problems. I will investigate the problems in turn to show that they cannot be resolved simply by making the tools fairer. Finally, I will suggest that Shelby's fairness criterion is similarly inadequate to justify punishment in the face of persisting systemic injustice.

1. Injustice is Inherent in Criminal Databases

The fundamental shortcoming of risk prediction under social injustice is that it relies on data shaped by a long history of injustice, both within the criminal justice system specifically and in society broadly. Risk prediction tools take these databases to be value-neutral in the calculation of risk scores, which simultaneously ignores and solidifies the injustice shaping them. As such, risk prediction tools make it impossible to intervene in the systemic causes of the correlation between crime data and socio-demographic variables. Justice requires that we identify and challenge such "causes of causes[14]" that risk assessment takes as a given.

In "Digitizing the Carceral State," Dorothy Roberts (2019) presents an overview of how the carceral state targets victims of injustice as a method of social control. Roberts argues that "despite claims that computerized prediction is objective, its databases and algorithms build in unequal social structures and ideologies that create new modes of state surveillance and control in marginalized communities" (1699). The technocratic discourses around big data, automation, and

[14] Geoffrey Rose, a 20th century epidemiologist, introduced the concept of "causes of causes" to public health scholarship. He argued, for example, that knowing lung cancer is caused by smoking is insufficient to address the problem of lung cancer. To do so, we must identify the factors, including social structural ones like poverty, that make people smoke. Similarly, knowing that people with criminal histories, for instance, are more likely to recidivate isn't enough to solve the problem of recidivism. We must understand factors like social injustice that affect the criminal histories of defendants.

prediction serve to hide these unequal structures and ideologies under the promise of scientific efficiency and accuracy. They do not, for instance, take into account that in recent decades, as the "federal, state, and local governments were dismantling the social safety net, they dramatically expanded their coercive functions, including increasing incarceration at unprecedented rates" (1700).

These policy choices about welfare and punishment are indicative of a trend in the US to "increasingly address social inequality by punishing the communities that are most marginalized by it." As such, they play a non-neutral role in shaping the databases that train AI risk assessment tools. For instance, in current US criminal law, "many of the basic conditions of being homeless" and "receipt of welfare benefits" are increasingly criminalized (Roberts 2019: 1703). "The school-to-prison pipeline is a well-documented pathway" in marginalized neighborhoods and is "especially perilous for black children" (1703). Many other factors related to social discrimination and marginalization, including addiction, mental illness, gender, race, sexuality, and disability, have also been shown to be increasingly criminalized (for an overview, see Brown and Schept, 2017). Given this differential treatment of advantaged and disadvantaged citizens by the carceral state, "past criminal history," a variable that increases a defendant's risk score across all variations of risk assessment in the US, disproportionately targets citizens who have been victimized by systemic injustice.

Crucially, systemic racial injustice affects risk factors beyond criminal history. This is why, even though race is not included in risk assessment tools, other variables come to act as "prox[ies] for race." In jurisdictions across the country, researchers have found that black communities and individuals disproportionately get identified as high risk (Roberts 2019: 1718). This is not necessarily because those risk predictions are inaccurate, as the ones described in the ProPublica article were, but rather because the databases and thus the resulting risk scores are

products of unjust policies. That black defendants are at higher risk for crime is an injustice created by the background structure of society. The risk predictions are thus objectionable on justice grounds, regardless of their accuracy. As long as the neighborhood, educational attainment, employment status, and criminal history of defendants are affected by unjust policies like residential segregation, employment discrimination, unequal public school funding, and racial policing, the risk scores will reflect that racial injustice. Getting rid of racially biased outcomes will not be enough to address this injustice.

These discriminatory policies reveal that data that are taken to be scientifically neutral predictors of risk in AI risk assessment tools are in fact shaped by severe injustice. Digitizing these socially and historically complex phenomena into quantitative risk scores overlooks this injustice and designates the individual defendant as the sole target of intervention. Developers of AI risk assessment tools aim for the most "accurate" prediction of recidivism without investigating the background structures that shape the trends in training datasets. This allows criminal justice actors to then justify using risk scores "according to a predetermined philosophy to punish instead of support marginalized communities" (1707). The status quo becomes treated as a given, and prediction becomes a method of replicating past injustices.

In the face of existing unjust background structures that shape the datasets of risk assessment tools, the popular pro-risk-prediction argument that computerized algorithms or statistical models are fairer than human decision-makers begins to crumble. As data that is shaped by unjust laws and institutions gets coded into the algorithms, the tools become an embodiment of the systemic bias they are purporting to escape. This form of bias is distinct from the bias in outcomes that is critiqued in the ProPublica article, and will persist even in the ideal "FAT algorithm." This is because, in the context of systemic social injustice, risk prediction quantifies injustice into a risk score that "necessarily reflects individuals' privileged or disadvantaged

positions" (1708). As such, while celebrated as reform, the AI risk assessment tool in fact exacerbates the existing background structure and thereby prevents progress towards justice.

Recognizing that injustice is coded into predictive algorithms reveals that resolving fairness problems like disparate error rates will not resolve the justice problem underlying crime and punishment data. While inaccurate predictions are certainly objectionable, they attribute injustice to the wrong source. Not only the outcomes but also the input of the tools must be unbiased for their use to be justified. Justice cannot be achieved by simply tinkering with the coding of the tools to produce more accurate or unbiased results, because the fundamental problem is not one of predictive accuracy or fairness, but rather of systemic injustice. Even if the tools produced the same error rates for different groups, and even if their predictions were perfectly accurate[15], their outcomes would be unjust.

2. AI Risk Assessment Tools Empower Developers, Disempower Criminal Justice Actors and Defendants

The second problem for AI risk assessment is that it empowers the developers of the tools at the cost of disempowering criminal justice actors and defendants. As their algorithms become increasingly complex and require expert knowledge, risk assessment tools empower an elite group of highly-educated coders and developers in determining the outcomes of marginalized offenders. In doing so, the digitized carceral state "concentrates administrative power in the hands of a small elite" (Eubanks, 2018: 200) who control the algorithms despite having little to no knowledge about the operations of the criminal justice system or the histories of the databases they are manipulating. This limits the ability of both criminal justice actors and defendants to acknowledge and challenge the complex trends of structural injustice that become reduced to simple risk scores.

[15] I will explore the dangers of prioritizing predictive accuracy under the third problem.

As structural injustice gets codified into the tools by this elite group of external actors, it becomes increasingly difficult for officials within the criminal justice system to see and challenge the logic underlying the tools. While some calculations of the algorithm behind risk scores may be beyond all human observation[16], key information like which variables are included in the calculations is nonetheless accessible to the developers of the tools. However, even such basic information may be unavailable to criminal justice actors, since the tools' algorithms tend to be proprietary[17]. Judges, who do not have access to and are unlikely to have expertise in the workings of the algorithm producing the risk score, are thus unable to challenge it. This diminishes judges' ability to rely on their discretion and make exceptions to the rules when risk scores obscure the nuances of individual cases or produce an outcome that may go against what justice requires.

Of course, judges may not always be moved to use their discretion to address injustice, and their discretionary judgments may be shaped by the very ideologies behind that injustice. Despite this, judges are nonetheless uniquely positioned to recognize and act on injustice in a way that algorithms are not, and often do use their discretion to protect marginalized defendants. In *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, Victoria Eubanks (2018) argues that past criminal justice reforms that minimized human actors' ability to appeal to their duty of justice compounded injustice. Eubanks points out that "minorities fare much worse under mandatory sentencing laws and guidelines than they did under a system

[16] This is the notorious "black box" problem of AI, where the output but not the decision-making process of the predictive tool becomes accessible to humans as the AI begins to learn on its own. This is a separate problem that would persist even if developers shared all of their knowledge with judges. I will not address it here, but it should be noted that it is one of the issues FAT algorithms aim to address.

[17] For instance, Northpointe, the company behind COMPAS, considers the tool's algorithms to be trade secrets and does not disclose how COMPAS determines individual risk scores or how it weighs various factors in arriving at a risk score to judges or defendants.

favoring judicial discretion. By depriving judges of the ultimate authority to impose just sentences, mandatory sentencing laws and guidelines put sentencing on auto-pilot" (81). As AI risk prediction tools increasingly put decision-making in all stages of criminal justice on auto pilot, injustices are likely to become further obscured and human actors unable to use their discretion to challenge them.

Eubanks' emphasis on the value of judicial discretion suggests that she shares Shelby's commitment to our collective duty of justice as a necessary tool for changing society's unjust background structure. While recognizing that state agents' judgments are often colored by racism, sexism, and disdain for poor people, she points out that these agents are open to change in a way algorithms are not. She writes, "I find the philosophy that sees human beings as unknowable black boxes and machines as transparent deeply troubling. It seems to me a worldview that surrenders any attempt at empathy and forecloses the possibility of ethical development" (2018: 168). She cites families who have interacted with child protective services and who would "rather have an imperfect person making decisions about their families rather than a flawless computer. 'You can teach people how you want to be treated'" (2018: 166-167).

Like the families mentioned above, criminal defendants whose destiny gets shaped by risk assessment tools increasingly lose the opportunity to appeal to the sense of justice of human actors. When their complex life narratives get reduced to what one researcher calls "the data fumes of human experience" (Barabas, 2019: 15), marginalized defendants become unable to tell their side of the story, which further obscures the injustice affecting their lives. The recent case of *State v. Loomis* exemplifies this phenomenon of disempowerment. In 2016, Eric Loomis filed a motion challenging the role COMPAS played in the determination of his prison sentence. Loomis argued that "the court's use of the COMPAS assessment infringed on both his right to an individualized sentence and his right to be sentenced on accurate information" (State v. Loomis

2017: 1531). The trial court denied Loomis' motion, and the Wisconsin Supreme Court affirmed the trial court's decision. On the accuracy claim, Justice Ann Walsh Bradley argued that COMPAS uses only publicly available data and data provided by the defendant, which Loomis could have denied or explained away. In making this claim, the justice apparently overlooked the fact that the ability of defendants to "explain away" what is presented as a scientific and objective prediction is likely to be severely compromised. On the individualized sentence claim, Justice Bradley "stressed the importance of individualized sentencing and admitted that COMPAS provides only aggregate data on recidivism risk for groups similar to the offender" (1532) but nonetheless denied the motion, arguing that COMPAS is only one component of the final sentencing decision and "courts have the discretion and information necessary to disagree with the assessment when appropriate." This argument ignores the fact that courts may not have the necessary information, such as the decision-making methods of the risk prediction tool, and that judges may consequently be hesitant to use their discretion against what appears to be an objective risk score. The claims raised in this case and its verdict reveal that, as risk scores gain authority in judicial decision-making, the nuances of defendants' stories and the injustices they have personally experienced start to lose their explanatory power. In the rare case when a defendant gets to speak up about this disempowerment, the courts appear to be uninterested in his justice claims.

Automated tools block our human capacity to recognize and fight injustice. They are, as such, a dangerous intervention to operate against an unjust background structure, regardless of their predictive accuracy. As Roberts argues, "human biases can be exposed, resisted, and potentially transformed, whereas computer algorithms cement biases into automated systems" (2019: 1712). AI tools disguise "overt discrimination based on demographics and socioeconomic status" (Starr 2014: 66) under the inscrutable, technocratic language of risk scores. When "words

yield to numbers" in criminal justice decisions in this way (Hamilton 2015: 13), the human claims of justice are more likely to go unheard.

3. "Accurate" Risk Prediction Perpetuates Injustice

The final justice problem for AI risk assessment tools concerns the fundamental philosophy of risk prediction guiding them. The success of risk prediction tools is measured on the tools' "predictive accuracy." This was, for example, the metric used in the ProPublica article to reveal racial bias in the COMPAS tool. The more accurately a tool can predict the future, the more it presumably improves the criminal justice system. However, given that predicting the future in the true sense is impossible, the focus on predictive accuracy leads to a conflation of replicating the past with predicting the future. This makes it all the more difficult to change existing unjust structures. Specifically, focusing on predictive accuracy solidifies existing injustices in four ways: First, it expands the surveillance power of the carceral state. Second, it promotes false positive errors, which are more difficult to identify than false negative errors. Third, it essentializes offenders and disregards their ability to change through state or other interventions. Fourth, it replicates and worsens past trends of injustice through a "self-fulfilling feedback loop" between prediction and punishment.

Predictive tools justify the expansion of the carceral state into all aspects of defendants' lives, since tools that use more variables are presumed to be more accurate. The more the tool "knows" about a defendant, the more precisely it can predict the defendant's future. Since there is no limit to the amount of data AI tools can sort through, data collection about any aspect of a defendant's life is possible, and can be justified by the goal of increasing predictive accuracy. There is thus no sphere of defendants' lives that is off limits to the criminal justice system. This expands the reach of the carceral state, where institutions outside of the criminal justice system become increasingly incorporated into a "carceral approach."

Roberts writes that, “under a carceral approach, the state’s aim is to control populations rather than adjudicate individual guilt or innocence, to manage social inequalities rather than to aid those who are suffering from them” (2019: 1712). Under this carceral approach, “institutions that provide medical, financial, labor market, and educational services have also become “surveilling institutions” for the carceral state, further blurring the line between social supports and carceral surveillance of poor populations and communities of color” (Barabas 2019: 9). Risk assessment tools justify expanding the reach of the carceral state to these “surveilling institutions” by equating more data with more accurate predictions and thus fairer punishment practices.

Second, focusing on predictive accuracy is likely to promote more punitive outcomes, because false positive errors are less likely to be recognized than false negative errors. A false positive error is said to have occurred when a defendant who is deemed high-risk does not commit a crime. A false negative error occurs, on the other hand, when a defendant deemed low-risk commits a crime. False positive errors are more difficult to detect because defendants who get classified as high-risk often end up incarcerated for extended amounts of time, and defendants classified as low-risk do not. It is impossible to know if a high-risk prisoner would have committed a crime had he been released into the community until the moment he is eventually released from prison. Predictions for high-risk offenders are thus often presumed accurate, since the defendant is not granted the freedom to enter society and prove the prediction wrong. There is more room for “error” in the case of low-risk defendants who are released into the community, simply because they actually have more opportunities to go against the prediction of the risk assessment tool.

When predictive accuracy becomes the measure of success in prediction, the unequal likelihood of false negatives and false positives will promote a punitive approach. To understand

why, simply consider this: A tool that gives life sentences to all defendants would be 100% accurate, as the possibility of false positive errors would be eliminated. This example reveals that, when accuracy becomes "a fetishized measure of a tool's worth" (Barabas 2019: 18), the normative value of accuracy goes uninvestigated. As Barabas puts it, for many proponents of risk prediction, "in cases of life and death, it doesn't matter *why* a prediction is accurate, so long as it is." This fetishizing perspective is captured well by the comments of a prominent statistician and criminologist: "I'm not trying to explain criminal behavior; I'm trying to forecast it. If shoe size or sunspots predicts that a person's going to commit a homicide I want to use that information, even if I have no idea why it works." (Barabas 2019: 19). Focusing on predictive accuracy alone is likely to incentivize higher risk predictions that lead to more punitive sentences. As such, the goal of predictive accuracy must be counterbalanced with a tolerance towards living with risk, and humility about our human inability to predict the future.

Third, risk prediction essentializes offenders and discounts the rehabilitative potential of state interventions. Focusing on predictive accuracy incentivizes criminal justice actors to treat risk scores as destiny rather than acknowledge the agency of defendants and promote their ability to act against the prediction through rehabilitative methods. The argument that people's future behavior can be predicted based on their past conduct implies that they are not worthy targets of rehabilitation. By codifying certain types of behavior or information as risk factors, the tools designate individuals who have a particular risk factor as a distinct kind of person from those who don't. This approach reduces individuals to their risk scores and ignores their unique needs, desires, and motivations. For instance, a risk prediction tool will treat all defendants with a particular criminal history the same way. It cannot account for the "causes of causes," such as the role state-perpetuated injustice, in how the individuals came to possess that criminal history.

This is not a new phenomenon. AI risk assessment is the latest iteration in a global trend where scientific categorizations are used to justify essentialism and discrimination. Roberts points out that "both eugenics and computerized predictive analytics rationalize continuing structural inequality by conflating forecasting the future with replicating the past. Thus, the predictive model that animates the contemporary carceral state has deep roots in US oppressive ideologies supported by mainstream science" (Roberts, 2019: 1714). In the US criminal justice system specifically, crime statistics have been used since the turn of the nineteenth century, when statistics about the overrepresentation of African Americans in prisons were used to justify arguments about the inherent criminality of that population and thus justify punitive policies towards them (Muhammad 2010: 46). Given this history, we should be hesitant to herald predictive accuracy as an objective, apolitical metric.

Fourth, and finally, targeting predictive accuracy alone is likely to perpetuate and worsen unjust practices by forming a self-fulfilling feedback loop where predictions become verified by the generation of more discriminatory arrests. (Ferguson, 2017: 513-515). This feedback loop forms when the "original selection bias arising from structural inequities generates observation bias which produces confirmation bias" (Roberts, 2019: 1721). As I pointed out when discussing the first problem, risk scores replicate past discriminatory trends by designating marginalized citizens as higher risk. Upon recognizing this disparity in risk scores, criminal justice actors may, either implicitly or explicitly, start to view those individuals as more dangerous. This may in turn make them more likely to be more punitive towards individual citizens coming from the same marginalized backgrounds, regardless of their actual risk profile.

For instance, a judge may come to see a defendant in a more negative light after associating him with similar defendants who received high risk scores, and consequently give him a longer sentence than he would have otherwise. Alternatively, law enforcement officials may

become more attentive to criminal activity in a neighborhood deemed "high-risk" and arrest more individuals than they would have otherwise. In both cases, the prediction appears to be confirmed, even though the prediction itself creates the circumstances for its confirmation. The racial profiling practices of the 1990s can be attributed to such a self-fulfilling feedback loop:

> Officials pointed to statistics that reflect the overrepresentation of African-Americans and Latinx in jails in order to justify racial profiling. They argued that their officers stopped and searched a disproportionate number of minorities, not because of racial animus, but because, quite simply the data showed that "that's where the criminals are." These officials used arrest and incarceration data as a substitute for crime rate and, in doing so, laid the foundation for the state's own recursive logic, whereby it used internally generated numbers about arrest and incarceration as a justification for continuing the very practices that fueled those numbers. (Barabas 2019: 31)

Associating certain groups of people with higher risk based on past data thus makes it more likely for people in those groups to get targeted in the future, thereby creating a feedback loop that targets historically marginalized people regardless of their risk profile. Prediction leads to confirmation in a chain of biased cognition.

Predicting risk is a dangerous endeavor, especially when predictive accuracy becomes the primary measure of success. Not only is focusing on predictive accuracy insufficient to address past and current injustices, but it is also likely to amplify them. It renders the "causes of causes" of crime irrelevant and disincentives efforts to go against the destiny written into risk scores. Society's failures to address injustice and prevent crime go ignored, and risk prediction appears fair, objective, and ethical as long as tools are deemed "accurate" (Roberts, 2019: 1724). This can make the privileged members of society, including policy and law-makers, rationalize their position of power and become oblivious to the injustice punishment continues to perpetuate towards marginalized offenders. Technocratic language downgrades the agency of offenders, and makes it difficult to see that crime is at times a reaction to social deprivation and a lack of lawful life choices. In a time when many in the US are oblivious to the persisting injustices of the

country, such tools can promote colorblind ideologies and blaming attitudes towards marginalized citizens. This rhetoric makes it more difficult for us all to recognize the injustices that surround us.

Revisiting Shelby's Fairness Criterion and the Promise of Just Punishment

A narrow focus on fairness in outcomes and the problematic metric of predictive accuracy it relies on make the promise of fair risk assessment an inadequate reform proposal to make punishment just. Taking a broader view of the systemic background injustices will allow us to reframe the limited discourse around risk prediction and address its shortcomings. In doing so, we must move beyond interpreting crime data as a way to categorize and control "risky" populations, and begin to challenge the assumptions of the carceral approach and the role of systemic injustice in producing and perpetuating crime statistics. As Kaya Williams argues, the philosophy of risk prediction exemplifies the American "captive imagination" (2017: 54), which traps people in "a logic of public safety and risk that casts the public as white, the risk as black, and safety as a sacred right of the white public to be violently protected no matter the cost. In such logics the acts of violence that expose public *un*safety are understood as the materialization of the latent threat inherent in black bodies rather than as the concrete result of the material conditions of life in the city and its jail."

Challenging the logics of risk and safety requires us to look beyond fairness in AI risk assessment. It is not just the shortcomings of the tool but rather its narrow placement in a broader system of injustice that makes AI risk assessment morally objectionable. Given this, simply making the tool fairer is not sufficient. We must strive to correct the background injustice, which the focus on "accurate" risk prediction can only serve to replicate.

In the same way, fair yet narrow criminal justice interventions cannot avoid injustice when society operates against severely unjust background conditions. Shelby's fairness

requirement is thus insufficient to make punishment just. In the context of AI risk assessment tools, it can only account for some of the problems outlined in previous sections, like the racially disparate outcomes of the tools. It cannot recognize the broader system of unjust policies and institutions that leave marginalized people at higher risk of committing crimes. This means that, even if the racial disparity in the predictive accuracy of the tools were resolved, the risk scores would nonetheless continue to be shaped by the background structure of society and thus remain objectionable on justice grounds.

Similarly, Shelby's requirement that criminal justice actors "be impartial and evenhanded in applying the rules" ignores the fact that the "rules" are not equally easy for all citizens to follow. For marginalized citizens who have been targeted by discriminatory laws and policies in healthcare, finance, education, and housing, adhering to the rules itself is an unjustly significant challenge. Just as a predictive tool will perpetuate injustice even as it delivers unbiased outcomes, a criminal justice apparatus that impartially applies the rules will disregard and thus leave unaddressed the fundamental background injustice of society. Fairness in one institution alone will not trickle down to the rest of society. As long as the background structure of society remains unjust, punishment will perpetuate further injustice to the marginalized members of society.

That Shelby's new framework fails to resolve the old dilemma of criminal justice under social injustice should not drive us to despair. To the contrary, it should push us to fight more diligently and urgently against injustice in all its iterations, and to not settle for narrow reforms that leave root causes of crime unaddressed. The US is one of the richest, most powerful democracies in the world. The state and its citizens collectively have the tools at their disposal to fundamentally reshape society's background structure.

Given that social justice is a necessary component of a just punishment scheme, our criminal justice reforms need to include social justice reforms. Emerging technologies like AI can be a part of this social justice mission. They need not bind us to perpetuating existing injustices; they can instead be co-opted as tools of radical change that disrupt the structures they operate in and on. Indeed, activists and academics alike have recently recognized this technology's potential to identify sources of institutional and historical injustice in ways previous methodologies cannot offer. Activist organizations like Data for Black Lives, the Black Futures Lab, and Our Data Bodies, for instance, use algorithmic analyses of government data to "reframe key debates around the use of data analytics for social justice" (Barabas 2019: 11).

When used for emancipatory ends, risk assessment tools would learn from the mistakes of current ones. AI risk assessment tools today disregard and thus reinforce unjust background structures, concentrate power in the hands of the elite, and treat prediction as destiny. An abolitionist perspective would use risk assessment tools to identify sources of injustice and targets of intervention, rather than taking the trends in data and the background injustice driving them as a given. It would ensure that marginalized citizens have a say in the development and application of the tools, so that the biases of elite developers don't transfer to the tools. Finally, as Roberts argues, abolitionist tools would "end prediction as a way of foreclosing social change by collapsing the future into past inequality" (Roberts, 2019: 1727). Instead, they would "facilitate envisioning a future that doesn't replicate the past." Rather than taking background conditions as a given and risk scores as destiny, an abolitionist approach would use such tools to develop reform and rehabilitation initiatives. For instance, algorithms could be developed to identify racially biased policies or help identify the needs of marginalized offenders and communities.

The social justice mission requires us to focus on minimizing rather than maximizing predictive accuracy. Instead of trying to eliminate the inherent uncertainty of risk models by

perpetuating feedback loops between prediction and confirmation, we need to embrace uncertainty "as enabling rather than disabling." This involves reframing "the nonexact coincidence between scientific predictions and observed reality as the promise, rather than the end, of science and of politics" (Chun, 2015: 678-679), since a just society would necessarily look radically different from one informed by past trends of injustice.

In other contexts where risk assessments yield undesirable projections, minimizing rather than maximizing predictive accuracy is already accepted as a desirable goal. For instance, rather than accept their projections about global warming as the inevitable doom of humanity, environmental scientists and activists use them to design policies that could prevent that prediction from becoming reality. The same can be done to change the behavior of individuals and man-made institutions that perpetuate injustice. Technology can be a tool to further the cause of justice and envision a radically different future, rather than a justification for the reproduction of historical injustices. In this radically different future where all citizens' constitutional essentials are met, state-imposed punishment would finally be free of moral taint.

Conclusion

The dilemma of criminal justice under conditions of social injustice cannot be resolved even if, as Shelby suggests, punishment is administered through a fair criminal justice apparatus, directed solely at *mala in se* crimes, and not expressive of moral condemnation. A criminal justice system that creates order through repression of urgent social problems cannot properly be called an institution of justice. For the repressive practice of punishment to be justified, society itself must be made so. Placing criminal justice within the framework of social justice in this way requires us to take a broader view of the systemic causes of crime. Crucially, as Angela Davis suggests, it requires us to explore "new terrains of justice, where the prison no longer serves as our major anchor" (Davis 2003: 21). Such new terrains would refocus our efforts on preventing crime rather than punishing it after it occurs. Society would then be made safer not by imprisoning the victims of injustice for their reaction to their life conditions, but rather by improving those life conditions through investment in increased social welfare, schools that offer social mobility, employment opportunities that provide a living wage, safe havens for people in danger of violence, and restorative policies that address past and current injustices across all social institutions.

Undoubtedly, the large, systemic changes that this philosophical reorientation requires will take a lot of time and political will. It will also require advantaged citizens to give up a lot of their privilege in the name of justice. However, the difficulty of the task cannot diminish its urgency. If we wish to rid ourselves of the moral taint of living in a society where the most marginalized citizens are punished for their reaction to their marginalization, we need to commit ourselves to social justice. We need to choose, as Bazelon (1975) calls us to, that we want our society to stand not on repressive order, but rather on moral order, no matter how long, painful,

and costly the process may be. Settling for a less ambitious goal would be a shameful act of injustice.

Works Cited

Angwin, Julia, et al. "Machine bias." *ProPublica, May* 23 (2016): 2016.

Baldwin, James. *The cross of redemption: Uncollected writings*. Vintage, 2011.

Barabas, Chelsea. "Beyond Bias: Re-imagining the Terms of "Ethical AI" in Criminal Law." *Criminal Law (April 25, 2019)* (2019).

Bazelon, David L. "The morality of the criminal law." *S. Cal. L. Rev.* 49 (1975): 385.

Benko, Jessica. "The Radical Humaneness of Norway's Halden Prison." *The New York Times*, The New York Times, 26 Mar. 2015, https://www.nytimes.com/2015/03/29/magazine/the-radical-humaneness-of-norways-halden-prison.html.

Brown, Michelle, and Judah Schept. "New abolition, criminology and a critical carceral studies." *Punishment & Society* 19.4 (2017): 440-462.

Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." *Conference on fairness, accountability and transparency*. 2018.

Chun, Wendy Hui Kyong. "On hypo-real models or global climate change: a challenge for the humanities." *Critical Inquiry* 41.3 (2015): 675-703.

"Criminal Law - Sentencing Guidelines - Wisconsin Supreme Court Requires Warning before Use of Algorithmic Risk Assessments in Sentencing - State v. Loomis 881 N.W.2d 749 (Wis. 2016)." *Harvard Law Review*, vol. 130, no. 5, March 2017, p. 1530-1537. *HeinOnline*, https://heinonline.org/HOL/P?h=hein.journals/hlr130&i=1552.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. "Extraneous factors in judicial decisions." *Proceedings of the National Academy of Sciences* 108.17 (2011): 6889-6892.

Davis, Angela Y. *Are prisons obsolete?*. Seven Stories Press, 2003.

Dreisinger, Baz. *Incarceration nations: A journey to justice in prisons around the world*. Other Press, LLC, 2016.

Duff, Antony, and Robert Alexander Duff. *Punishment, communication, and community*. Oxford University Press, USA, 2001.

Eubanks, Virginia. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

Ferguson, Andrew Guthrie. "Illuminating Black Data Policing." *Ohio St. J. Crim. L.* 15 (2017): 503.

Garrett, Brandon L. and John Monahan. “Judging Risk” (July 20, 2018). *California Law Review*, Forthcoming; Virginia Public Law and Legal Theory Research Paper No. 2018-44. Available at SSRN: https://ssrn.com/abstract=3190403

Hamilton, Melissa. "Adventures in risk: predicting violent and sexual recidivism in sentencing law." *Ariz. St. LJ* 47 (2015): 1-62.

Kelly, Erin I. *The limits of blame: Rethinking punishment and responsibility*. Harvard University Press, 2018.

King, M.L., Jr. (1963). *Letter from The Birmingham City Jail*. This version, originally released without copyright in 1963 and disseminated widely, is assumed to be in the Public Domain.

Lewis, Christopher. "Inequality, incentives, criminality, and blame." *Legal Theory* 22.2 (2016): 153-180.

Lum, Kristian, and James Johndrow. "A statistical framework for fair predictive algorithms." *arXiv preprint arXiv:1610.08077*(2016).

Muhammad, Khalil Gibran. *The condemnation of blackness: Race, crime, and the making of modern urban America*. Harvard University Press, 2010.

Roberts, Dorothy E. "Book Review, Digitizing the Carceral State." *Harv. L. Rev.* 132 (2019): 1695-1708.

Shelby, Tommie. (2016). *Dark ghettos: Injustice, dissent, and reform*. Harvard University Press.

Starr, Sonja B. "Evidence-based sentencing and the scientific rationalization of discrimination." *Stan. L. Rev.* 66 (2014): 803.

Tadros, Victor. "Poverty and criminal responsibility." *The Journal of Value Inquiry* 43.3 (2009): 391-413.

Watson, Gary. "A moral predicament in the criminal law." *Inquiry* 58.2 (2015): 168-188.

Williams, Kaya Naomi. "Public, Safety, Risk." *Social Justice* 44.1 (2017): 36-61.