Philosophy & Technology
https://doi.org/10.1007/s13347-021-00491-2

RESEARCH ARTICLE

Abolish! Against the Use of Risk Assessment Algorithms at Sentencing in the US Criminal Justice System

Katia Schwerzmann[1]

Received: 9 November 2020 / Accepted: 2 November 2021

Abstract
In this article, I show why it is necessary to abolish the use of predictive algorithms in the US criminal justice system at sentencing. After presenting the functioning of these algorithms in their context of emergence, I offer three arguments to demonstrate why their abolition is imperative. First, I show that sentencing based on predictive algorithms induces a process of rewriting the temporality of the judged individual, flattening their life into a present inescapably doomed by its past. Second, I demonstrate that recursive processes, comprising predictive algorithms and the decisions based on their predictions, systematically suppress outliers and progressively transform reality to match predictions. In my third and final argument, I show that decisions made on the basis of predictive algorithms actively perform a biopolitical understanding of justice as management and modulation of risks. In such a framework, justice becomes a means to maintain a perverse social homeostasis that systematically exposes disenfranchised Black and Brown populations to risk.

Keywords Predictive algorithms · Cybernetics · Criminal justice · Social justice · Biopolitics

1 Introduction

The idea of a society governed by predictive algorithms has become a common foundational trope of mainstream cultural productions within the past 20 years. In the film *Minority Report* (2002), the job of John Anderton—the main protagonist played by Tom Cruise—is to arrest people before they commit a murder. While the crime is pre-viewed by the *Precogs*—prescient human beings kept in a state of artificial coma—John's job is to interpret the images extracted from their brains to determine where the murders will be committed. The predictive dispositive in *Minority Report* is characterized by a mix of enhanced human ability with technologies of

✉ Katia Schwerzmann
katia.schwerzmann@uni-weimar.de

[1] Media Studies Department, Bauhaus Universität Weimar, Weimar, Germany

extraction and the recording of information. Danny Witwer, whose job is to audit the Precrime Unit and evaluate its procedure, asks John: "But it's not the future if you stop it. Isn't that a fundamental paradox?" In response, John throws him a wooden ball on which the name of a perpetrator had been imprinted. As Danny catches the ball before it falls, John answers: "The fact that you prevented it from happening does not change the fact that it was going to happen."

More recently, the series *Westworld* (2016–) grapples with the relation between artificial intelligence and issues of self-determination, memory, and the exploitation of A.I.-endowed humanoids by humans during violent roleplays. The exploitive relationship between humans and machines gets overturned when one of the humanoid machines—Dolores—kills her creator during a mass shooting perpetrated by the machines against humans. In the third season of *Westworld* (2020), Dolores is attacked and badly wounded. She is saved by Caleb, a socially outcast human. Dolores explains to Caleb that since it has been algorithmically predicted that he will end up killing himself, he is not deemed worthy of any social investment: "They won't invest in someone who's going to kill himself. But by not investing, they ensure the outcome." The prediction of Caleb's fate triggers a feedback loop that might well achieve the prediction.

From *Minority Report* to *Westworld*, one notices a shift from a deterministic toward a probabilistic model of the world. While predictions in the first model accomplish themselves because the future is known by creatures with the godlike ability of foresight, predictions are realized in the probabilistic model through a complex process of data-gathering and processing, and decision-making based on the algorithmically produced predictions. However, their results are the same: the predicted future is inescapable.

The above cultural productions depict neo-liberal society as characterized by its exclusive focus on risk management, a process by which every course of action is evaluated in terms of calculated risk and its return on investment. Subsequently, the result of such evaluations is directly fed back into the processes of decision-making. Risks are "calculated" because neo-liberal capitalist societies do not seek to exclude risk per se. Rather, they rely on using risk as a motor for carefully planned change as a mode of governance. As Stefano Harney and Fred Moten insist in *The Undercommons*, this mode of governance submits the population—primarily the disenfranchised and racialized portion of it—to increasingly higher levels of contingency, flexibility, and thus risk (Harney & Moten, 2013, pp. 76–79). This deliberate instability, which weakens social ties, renders individuals susceptible to ongoing adjustment and control. Following the recursive logic of the regulation and control of society through risk management tools, the future escapes the program less and less; its openness diminishes with every optimization.

1.1 Literature Review

For the past two decades, predictive algorithms called "risk assessment tools" have been widely used by state courts of the US criminal justice system for judicial decision-making in regard to a defendant's access to rehabilitative programs, and the

granting of probation and parole. More recently, they have been used to help determine the sentence of an offender. The reason advanced by the proponents of these tools is that they "reduce recidivism and increase public safety" (State of Wisconsin v. Eric L. Loomis, 2016, §2). Justifying their use, they argue that algorithmic tools would achieve this by being more transparent and accurate in their prediction of recidivism than a judge's assessment.

Both the objectivity and efficiency of these algorithmic tools have recently come under scrutiny in many studies—among which a broadly discussed inquiry by ProPublica (Angwin et al., 2016). Existing critiques of predictive tools generally focus on issues of lack of transparency and the biases ingrained into the algorithms and data they are trained on and which reflect firmly rooted societal biases (Noble, 2018). One of the main arguments in this broader public discourse is that algorithms conceal that they are the product of human decision-making at every stage of the process of their production and that the data they are trained on reflect the racist and classist biases of the society that produces them. This leads to the naturalization of contested categories, foreclosing them to political discussions (Eubanks, 2018) (O'Neil, 2016) (Benjamin, 2019). The black box character of algorithms, which is in important part caused by their proprietary nature, makes their functioning opaque and inaccessible to the public (Brauneis & Goodman, 2018). Following Campolo and Crawford, the conjunction of the accuracy of predictions with the inaccessibility and the opacity of predictive algorithms grants these algorithms a magical power or "enchanted determinism" that allows their producers and users to remove themselves from the responsibility tied to their decisions (Campolo & Crawford, 2020).

On the other hand, a significant amount of research is invested into questioning these critiques and justifying the use of predictive algorithms. This research seeks to compare the human ability to make accurate predictions to its algorithmic counterparts (Lin et al., 2020). Discussions revolve around the necessary tradeoff between fairness and public safety (Barabas et al., 2017). From this perspective, the problem of accuracy and biases is overcome through technical adjustments in the computational model: a better adjustment of the tool should allow the resolution accuracy issues and minimize biases (Corbett-Davies et al., 2017). The question of fairness is posed in statistical terms. Fairness becomes a question of tradeoff between the equality in treatment between racial groups and considerations around public safety. Depending on how fairness is defined, one can argue that the use of algorithms is fair and even fairer than human decision. Recently, Wong (2020) has appropriately argued that in a democracy, the definition of fairness should be the product of political decision, rather than be left to statistical considerations. These statistical considerations around fairness are justified by a consequentialist understanding of justice (Card & Smith, 2020). From a consequentialist standpoint, risk assessment algorithms appear as adequate and efficient means to an end, which consists in assuring public safety while better managing limited public resources. I do not argue from a consequentialist perspective because I consider it part and parcel of the issues I subsequently address. In addition, I understand the consequentialist framework as computational at its core in that it consists in the calculus of the means toward an end deemed "good." The issue I see with this approach is that no calculus can ever ground this "good." As a norm, the "good"—in the case of justice, "fairness"—is

the object of an always-contestable decision—here, for instance, between "public safety" and "racial fairness."[1] What a society decides is fair can never be fully grounded through calculation and must instead be instituted (Benjamin, 1996). The norm in any consequentialist and utilitarian framework is a blind spot that escapes the paradigm of the calculus of the means. Because this norm is itself the object of a contestable decision, any decision oriented toward it entails an undecidable or incomputable. By understanding decision-making in terms of pure calculation, the consequentialist framework implies a lack of perception toward the incomputable comprising any decision. If justice can never be the sole product of a calculus due to the ungroundable character of fairness, then the utilitarian, consequentialist framework hinders the discussion of issues tied to algorithmic-aided decision-making. Instead, it is necessary to move beyond this framework.

1.2 Arguments

While fully acknowledging the importance of approaches criticizing the racially biased character of algorithms and their lack of transparency, my line of inquiry differs significantly from the approaches above. Matters of transparency and bias, while important, leave other fundamental philosophical issues unchallenged; this is what I seek to tackle in this paper. These issues broadly regard the rewriting of the future induced by predictive tools tied to their recursive logic and the implicit understanding of justice as risk management. Following Gregory Bateson, I understand "recursivity" as the regulating logic by which the informational result of an algorithmic process is fed back into the system and influences the subsequent process (Bateson, 1979, pp. 126–127). The algorithmic rewriting of the future results in the exclusion of unprogrammed futures in favor of a future normatively deemed right or good. This "good" becomes all the more unquestionable as it seems to "naturally" result from algorithmic computation.

In this paper, I show that predictive tools cannot be salvaged for justice purposes and call for the abolition of risk assessment algorithms. In the Black radical tradition that informs my argument, abolition is not simply about repealing or destroying something; it is constructive and transformative.[2] Why then call for the abolition of what appears to only be a tool for judicial decision-making? As I argue based on the work of Karen Barad, N. Katherine Hayles and on Media Studies, a tool is never solely a tool, that is, an external, independent appendage that can be added to or removed from a system. Rather, technology emerges from and is conditioned by a

[1] A whole other paper would need to address what is considered as "public safety" here and whose interests and lives are prioritized in these considerations around safety. A lot of implicit norms are at play that are not solvable through computation.

[2] I am following here the call for abolition formulated by Davis (2005, pp. 72–73). And then reformulated by Stefano Harney and Fred Moten in Harney and Moten (2013, p. 42): "What is, so to speak, the object of abolition? Not so much the abolition of prisons but the abolition of a society that could have prisons, that could have slavery, that could have the wage, and therefore not abolition as the elimination of anything but abolition as the founding of a new society.".

society that it, in turn, transforms.[3] The logic that predictive algorithms operationalize, and which I characterize as recursive, have existed in western society since the rise of biopolitics in the nineteenth century. However, this logic is accentuated and naturalized through the apparent objectivity of algorithms. In that sense, abolishing the use of predictive tools goes hand-in-hand with questioning the conception of justice as risk management. In short, calling for the abolition of a "tool" means questioning the socio-technological dispositive without which there is no such tool. Repealing a dispositive never just means repealing the concrete material stuff of which the dispositive is made, but also simultaneously repealing its implicit theoretical framework. It implies rethinking the framework so that the material dispositive is not justifiable anymore. If one agrees to repealing predictive algorithms for the reasons I discuss in this paper, this necessarily entails questioning the reduction of criminal justice to the management of risks.

In what follows, I will outline three arguments to demonstrate that predictive tools should be abolished. My first argument regards who and what is judged when algorithmic tools constitute the basis for sentencing. Some researchers argue that the use of algorithmic tools for the purposes of justice leads to highly individualized evaluation thanks to the breadth of gathered data and other factors taken into account. Others consider that these evaluations are problematic because of their generalized character. I will show that the issue is not one of generalization vs. particularization. Instead, it is about the rewriting of temporality induced by the use of predictive tools.

Second, I will discuss how the use of algorithms transforms a probabilistic vision of the future into a deterministic one thanks to the decisional, and thus performative, character of justice. I mean performativity in its classical Austinian sense: the effect that a speech-act has on reality (Austin, 1962). Sentencing as a decision is an obvious speech-act, as it affects a very concrete change in the life of the sentenced individual as well as their family and community. Judicial procedures function as techno-political practices, in which the material conditions of computation and the performative power of decision are inextricably entangled. Algorithms alone neither describe nor determine the future. Rather, I argue, decisions based on algorithmic predictions create a reality that progressively confirms existing predictions—leading to a de facto determinism that has nothing to do with algorithmic omniscience. I will demonstrate that feedback processes, far from correcting predictions and thus decisions based on them, systematically suppress the outliers, transforming reality to match predictions.

In my third and final argument, I will show that decision-making premised on predictive algorithms repeatedly performs a biopolitical understanding of justice as the management and modulation of risks. Justice becomes a means to maintain a

[3] I rely here on three similar ways of thinking about this co-shaping relationship. Recent German media theory speaks of "cultural technique" (Siegert, 2013). The media philosopher and historian of cybernetics N. Catherine Hayles speaks of "technogenesis" (Hayles 2012). Building on Niels Bohr's thought experiments in quantum physics, Karen Barad describes this co-shaping as "entanglement" and "intra-action" (Barad, 2007).

perverse homeostasis—understood here as the conservation of a given state of the system—that implies systematically exposing disenfranchised Black and Brown populations to deadly risks. This final argument is based on a rereading of Michel Foucault's understanding of biopolitics as a specific kind of power exerted by the western states since the nineteenth century. Unlike sovereign power, biopower does not wield upon the individual body as much as it aims to manage and regulate a population in its entirety, its birth and death rate, and its health (Foucault, 2003). Instead, I aim to show that the risk-management framework of the criminal justice system should be understood not from the perspective of an entire population's management, but from the perspective of which population is systematically exposed to risks.

The performance of justice that comes to light through my three arguments is in conflict with a democratic understanding of justice where, following Jacques Derrida, the act of justice has to "always concern singularity, individuals, irreplaceable groups and lives, the other or myself as other, in a unique situation" (Derrida, 1990, p. 949). I base my argument on Derrida's definition of justice because it is a classical definition similar to the one found in the *Institutes of Justinian*—one of the foundational texts of the western legal system.[4] Moreover, I regard any definition that does not have the singularity and irreplaceability of the judged individual at its core as incompatible with democracy.

Algorithmic justice does not serve retribution or rehabilitation. Instead, it reveals itself as a mechanism of biopolitical regulation and control. My three arguments shall lay bare that correcting and optimizing predictive algorithms will not solve the issues brought to light in this paper because these corrective mechanisms do not question the problematic framework of risk management itself.

To substantiate my arguments, it is essential to first understand the functioning of these algorithms in their context of emergence. I will focus on COMPAS (by Equivant, formerly known as Northpointe Inc.) as a case study, as it is one of the most widely used predictive algorithms in the US criminal justice system.

2 COMPAS (Correctional Offender Management Profile for Alternative Sanctions) in Context

Risk assessment tools are not a new phenomenon. Actuarial tools have been in use in the US justice system since the 1960s. Among older predictive tools was the "Salient Factor Score" used from 1973 until the 1990s (Hoffman & Adelberg, 1980). However, the scope of the tool was limited as it provided an evaluation based on only 7 factors (against 137 with COMPAS) (Hoffman & Adelberg, 1980, p. 49). Factors on which the offender has little influence—like education, employment, family, and even their age at the first offense—were progressively removed from the evaluation in later versions of the tool (Tonry, 2014, p. 168).

[4] Cf.: "Justice is the constant and perpetual wish to render to everyone their due" (Holland, 1881, p. 4).

Predictive algorithms like COMPAS belong to the 4th generation of such tools (Brennan et al., 2009, p. 21; Hamilton, 2014, p. 238). COMPAS is one of the most widely used risk assessment algorithms in US courts. These tools are generally proprietary, so the public does not have access to their content or their exact functioning. While the goal of the "Salient Factor Score" was to predict the chances of success of rehabilitative measures, COMPAS focuses on the evaluation of an offender's recidivism risk as well as their specific "needs" (social, psychological, etc.) for rehabilitative purposes. Unlike previous tools like the "Salient Factor Score," COMPAS is used for sentencing. As a consequence, an individual who scores high on the risk of recidivism portion will likely spend more time in prison than if they were judged solely based on their offense.

Tools like COMPAS were designed in the context of a new conception of judicial practices called evidence-based practices (EBP). The declared goal of the EBP approach is to increase the financial efficiency of the criminal justice system while decreasing the prison population by more efficiently evaluating who should be granted parole or probation. EBP would also help determine what rehabilitative measures would be efficient for which kind of offenders (Couzens, 2011, p. 2). Following Cecelia Klingele, the EBP approach established itself after years of a punitive conception of justice that rejected rehabilitative measures as inefficient and, from the 1970s on, had led to a massive increase in the prison population (Klingele, 2016, p. 540). The specificity of EBP's methodology is that it relies on a statistical approach to judicial procedures.

It is worth taking a closer look at the way the COMPAS tool presents its results. It produces a report, which consists of two sections (cf. Appendix). The upper section "Overall Risk Potential" contains the evaluation of the overall recidivism risk ("recidivism," "violence," "failure to appear," "community compliance") on a scale of 1 to 10: 1 being the lowest risk, 10 the highest.[5] The lower section of the report titled "Criminogenic and Needs Profile" attributes scores to factors or "predictors" said to influence the risk of recidivism. There are two kinds of predictors: the static ones and the dynamic ones. Static predictors are those that belong to the history of the offender and cannot be changed: age, gender, and race fall into such predictors. As predictors like education, socioeconomic status, substance abuse, and social adjustment problems are susceptible to change and are potential targets of rehabilitative measures, they are considered dynamic (Hamilton, 2014, p. 237). The assessment based on these factors in the second section of the COMPAS report extends from the history of previous criminal involvement to "substance abuse," "social environment," "criminal thinking," criminality in the family, "socialization failure," "social isolation," and so on (see Appendix). While the first section was "developed using methods and strategies for predictive modeling" and enables "discriminat[ion] between offenders who will and will not recidivate" (emphasis mine), the scores in the second section "are not meant to be predictive but aim simply and accurately

[5] It is to note that the assessment retrieved by ProPublica and shown in the Appendix is dated 2006. The categories have slightly evolved since then as can be seen in the COMPAS Practitioner's Guide by Northpointe Inc. (2015).

to describe the offender (Northpointe Inc., 2015, p. 7)." (Note Northpointe's use of the future tense, which is significant for my argument.) The scores in both sections result from comparing the data of the offender with the data of the "norm group." Even if the scores of the second section are not meant to be predictive, the data tied to these scores contribute to determining the overall risk of recidivism.

Data on the individual are "gathered from the offender's criminal file and an interview with the defendant" (State of Wisconsin v. Eric L. Loomis, 2016, §13) during which the offender must answer the 137 questions on the "COMPAS Core" questionnaire. One of these questionnaires has been retrieved by ProPublica and is available online.[6] The scope of the questions used to assess the defendant is extremely broad. The individual is not only evaluated based on their history but on factors about which they have no or very limited influence like age, criminality in the family, socioeconomic background, or criminality in the neighborhood, among others. The answers of the assessed individual are then compared to the score of the "norm group." The dataset of the norm group consists of over "30,000 COMPAS Core assessments conducted between January 2004 and November 2005 at prison, parole, jail, and probation sites across the United States."[7]

As Northpointe *Practitioner's Guide* shows, COMPAS associates the assessed individual to a type of criminal. COMPAS distinguishes between eight categories of male criminals and eight categories of female criminals (Northpointe Inc., 2015, pp. 50–57). The algorithm classifies each offender into a typical profile. Northpointe points out that no offender ever fully matches their class. However, the classification supposedly helps "treatment staff conceptualize the 'kind' of client they are dealing with" (Northpointe Inc., 2015, p. 47). Such categorizations are not a novelty. Typologies of criminals were created as early as the 1930s (Harcourt, 2007, p. 180 f.). Far from being objective, these typologies are influenced by the values of the society from which they originate. 1930's typologies were shaped by moral values and were criticized for their arbitrariness (Harcourt, 2007, p. 181). COMPAS categories, on the other hand, bear the traces of the 1980s war on drugs and the criminalization of addiction; the first category of male offenders is "Chronic drug abusers – most non-violent" (Northpointe Inc., 2015, p. 50).

While COMPAS is gender-sensitive—which can be considered further problematic since individuals are judged differently based on a factor for which they bear no responsibility (Tonry, 2014, p. 171)—Northpointe's handbook does not mention race as a factor influencing the risk of recidivism score. However, this does not mean—and by a longshot—that race is not an integral part of the predictions. It has been shown that factors like neighborhood, socioeconomic status, e.g., which are used as predictors in

[6] See https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE, last accessed May 10, 2021.

[7] Northpointe Inc. (2015, p. 11): "The Composite Norm Group consists of assessments from state prisons and parole agencies (33.8%); jails (13.6%); and probation agencies (52.6%). The Core Norm includes 7,381 offenders. Men represent 76.9% of the Core Norm Group (n = 5,681), and women represent 23.1% of the Core Norm Group (n = 1,700). The median age at assessment is 31.0 (M = 32.6) in the Core Norm Group. The racial composition of the Core Norm Group is 61.6% Caucasian, 24.9% Black, 10.3% Latino and 3.2% other racial groups.".

COMPAS to evaluate recidivism risk, function as proxies for racial categories because of the still largely segregated American society (O'Neil, 2016; Mbadiwe, 2018, p. 19). Therefore, race does not need to be explicitly stated in the COMPAS Core questionnaire to influence the risk score of an individual by proxy. In short, no algorithm that takes socioeconomic factors and social environment into account is race-neutral.

From an evidence-based perspective, it is assumed that thanks to such an "objective statistical assessment" (Northpointe Inc., 2015, p. 3) tool, justice would become fairer and less prone to prejudice and bias (especially racial ones) by providing an "objective" evaluation of the offender. With an accuracy rate of 65%, a 2018 study by Dressel and Farid shows that COMPAS is not better at predicting the recidivism risk of an offender any more than a nonexpert participant asked to do the same when provided with "a short description of a defendant that included the defendant's sex, age, and previous criminal history, but not their race" (Dressel & Farid, 2018, p. 1), while in comparison, COMPAS bases its predictions on 137 features (Dressel & Farid, 2018, p. 2). Furthermore, in comparing the rate of false negatives and false positives, the study shows that human participants and COMPAS results "are similarly unfair to black defendants" (Dressel & Farid, 2018, p. 2). Other researchers have more recently argued that algorithms are more accurate than humans depending on the circumstances and recommend to include even more risk factors in the hope of increasing the accuracy of algorithmic predictions (Lin et al., 2020). Meanwhile, the US criminal justice system is gathering a vast amount of data that has little to nothing to do with the offenses for which an individual is judged. Seen in this light, risk assessment tools appear as instruments of massive control applied to every aspect of an individual's life.

With the evidence of such questionable results, one might wonder how these tools could become so essential to judicial procedures. The aura of efficiency and objectivity associated with risk assessment tools appears to have made decision-makers oblivious to the evidence that shows otherwise. For instance, while the State of Wisconsin acknowledges the limitations of risk assessment tools in a 2016 court decision, it still maintains that the court will make use of these tools for sentencing, referring to the aforementioned evidence-based practices (State of Wisconsin v. Eric L. Loomis, 2016, §66).

Here, I want to reemphasize the limitations of the research that focusses on the measurement of fairness and accuracy in algorithms. Measurements might show that algorithms are doing a slightly better job than judges, which, from an EBP (consequentialist) standpoint, should suffice to justify their use. However, such a discussion presupposes that we accept to reduce justice to the management of risks and resources, and that we equate judicial decision-making with computation. It is this conception of justice that I question in order to demonstrate its problematic character.

3 The Performativity of Predictions

In what follows, I will bring us a step further than the often-invoked biased character of risk assessment tools toward minorities. Again, algorithmic biases are a fundamental, yet not surprising issue given that these algorithms are trained on datasets

that are the products of a society suffering from structural racism. The point I would like to make is not that algorithms like COMPAS could or should be improved or optimized; improvement procedures presuppose the reduction of decision to computation, an assumption challenge in my third argument. Rather, their use at sentencing should be abolished.

3.1 Rewriting the Temporality of an Individual's Life

It is considered self-evident in the US criminal justice system that individuals are not judged solely on the offense that brought them before the judge, but that the history of past offenses is taken into account at sentencing (Tonry, 2014, p. 172). One example of this is the "three-strikes law" applied in 30 out of 50 US states. Regardless of whether the offender was already punished with prison time for the first two offenses, by the third strike and in the case of a violent or serious offense, the offender faces a life sentence. This, however, should be viewed as far from evident, considering that the offender would have already been punished for their past crime. Criminologist Michael Tonry notes in a 2014 paper that the Scandinavian countries have a very different conception of punishment, as they deem that past offenses for which the offender has already been punished should not be taken into account at sentencing (Tonry, 2014, p. 172).

A new factor comes into play with predictive tools, however: the risk of future recidivism. This risk is evaluated based on the defendant's criminal history but also on their overall recidivism risk in conjunction with the "class" of criminals to which the individual is attributed by the risk assessment tool. The data used to produce this evaluation do not solely belong to the individual. The predictors mentioned in "Sect. 2" are predictive only by comparing the individual's data with the data of the norm group. As a result, the individual is judged relative to the category of criminal to which they are expected to belong. Generalization through datafication is problematic in regard to the conception of justice I mentioned earlier, which has to address the singularity of each case. If it does not, justice becomes nothing more than the automated application of general rules, no matter how different singular cases are from each other.

However, the question of the generalization of sentencing through datafication is not the one I would like to ask here. In fact, both individualization and generalization happen in the assessment produced by predictive algorithms. Indeed, the algorithm treats the offender in a highly individualized way when it comes to the amount and specificity of data gathered regarding that individual's past[8]; on the other hand, it generalizes the data by simplification when it pertains to their future. As a result, the individual is judged relative to the category of criminal to which they are expected to belong—"expected" in the sense that predictions are probabilistic and do not amount to determinism. Probabilities only establish the frequency of an event occurring when another event takes place. For this reason, predictions do not

[8] Sociologist Bernard Harcourt questions the idea that the use of algorithms and statistics leads to a generalization away from the individual. On the contrary, he states: "The actuarial is better understood, instead, as the culmination or the zenith of the turn to the individualization of punishment" (Harcourt, 2007, p. 110). Harcourt claims that the goal of risk assessment tools is not to generalize but to indi-

establish a causal, necessary relation between both events. Producing a deterministic evaluation of the future of an individual would require—as is *Minority Report*—deity-like prescience or a system that disposes of the knowledge of the integrality of the factors and causal intricacies at a given stage of an individual's life on a quantum level.[9] The algorithm necessary to compute the history of this individual would have to be exactly as long and complex as this history itself. It would thus be useless. Following the concept coined by the mathematician Gregory Chaitin, the individual's history is incompressible: "… if the experimental data cannot be compressed, if the smallest program for calculating it is just as large as it is …, then the data is lawless, unstructured, patternless …. In a word, random, irreducible!" (Chaitin, 2005, p. 64). To predict the future of an individual necessitates discovering a pattern, and in order to do so, one must compare the data of the said individual to a dataset and thus give up their singularity. One has to trade off the certainty of an impossible determinism for the uncertainty of predictions. Patterning always already implies a simplification per generalization and, thus, a loss of certainty.

The question arising from this assessment is what allows a judge to act as if the future of an individual had already been lived. Indeed, no matter how many factors are included and processed by the predictive algorithm, the sentencing based on its result consists in judging a future that cannot be the future of the judged individual as this singular future has yet to be lived. This future is open, undetermined. It has not happened yet. Probabilities of recidivism can be high, but they are just that: probabilities. To base a decision on a probable outcome which is the outcome of a class of individuals means to deny an individual the openness of their future, in other words, to deny the multiplicity of possible outcomes, while implicitly devolving to the individual the whole responsibility for the social conditions in which they grew up, these conditions being used against them when gathered for prediction purposes. Incidentally, judging an individual based on a future that cannot be theirs can only be justified if justice is implicitly understood as the management of risks, while the present offense and the judged individual are secondary matters.

The denial of the indeterminacy of the individual's future causes the temporality of the individual's life to flatten into a present inescapably doomed by its past: the individual's past is used to predict the future as if this future had already been lived, and this "as if" serves in turn to performatively determine their present through judicial decision-making. The point here is that the issue lies not so much in the existence of statistics and predictions; rather, it consists in the practice of basing

Footnote 8 (continued)
vidualize punishment as much as possible by taking the highest possible number of factors into account. However, as sociologist Katherine Beckett shows in the late 1990s, this use of statistics does not serve a better understanding of what causes an individual's trajectory. Instead, it leads to disregard for the societal causes of crime (Beckett, 1999, p. 102). See also Klingele (2016, p. 574). With COMPAS, socioeconomic factors are not taken into account to better comprehend what could lead a specific individual to commit a specific crime, resulting in a more lenient sentence. Rather, used to produce a prediction, these factors serve to show that an individual will most likely not escape their socioeconomic background.

[9] This is by the way the scenario of a recent series, *Devs* (2020), by the director of the film *Ex Machina*, Alex Garland.

decisions on them for purposes of justice. The actual outcome—that the individual reoffends or does not reoffend—might match the predicted outcome, as in the case of an offender who was granted parole based on a low-risk score of recidivism and who does not reoffend. But in the case of a sentence to prison based on predictions of recidivism, there is no way to know what the actual outcome for this individual would have been. With such decisions, the future of this individual has been performatively determined as lived—the "he will reoffend" in *Northpointe Practitioner's Guide*—before it could actually be lived. For this reason, it is highly problematic to ground decisions about parole, probation, and time spent in a prison cell not on the present state and needs of an individual but on predictions about a future that cannot belong to the judged individual.

However, from the perspective of risk management, one would argue that society has both the right and good reason to protect itself from offenders by using all the knowledge and data at its disposal and that it is safer to falsely give a longer sentence to someone based on his or her risk of recidivism than freeing an individual by error who then happens to reoffend. This argument does not hold, as it has been proven that imprisonment does not make society safer: American prisons are criminogenic and imprisonment is highly detrimental to communities as it damages their social fabric (Clear, 2008).[10] What Angela Y. Davis calls the "prison-industrial-complex" (Davis, 2005, p. 35) incarcerates Black people at a much higher rate than their white counterparts for similar offenses, thus systematically overexposing this population to the risks tied to prison.

3.2 Probabilities vs. Decisions

In this section, I will discuss how the use of algorithms transforms a probabilistic vision of the future into a deterministic one thanks to the decisional and thus performative character of justice. Because decisions are performative, probabilities become deterministic: they produce the world they predict. To show this, I would like to turn to a discussion regarding the specific performativity of decision-making based on predictions and analyze what changes occur in the reality beyond the ones affecting the judged individual.

By taking the recidivism risk of an individual into account for sentencing, we are determining the future of this individual based, as we have seen in the previous section, on predictions tied to data that are not exclusively theirs. Because decisions are performative, by choosing between one or another outcome (prison or probation), we take probabilities as if they were deterministic: a given individual is predicted to recidivate and will thus be sentenced to prison as if they actually had reoffended. This decision supposes the existence of something, the future reoffense, that is not and cannot be, as the individual is now in prison.

[10] As Michel Foucault demonstrated in *Discipline & Punish* (Foucault, 2012), this critique against prisons is as old as prisons themselves and is fully part of the constant reformation process of the prison complex.

Earlier, I called attention to the striking formulation by Northpointe that I repeat here: "The purpose of the risk scales is prediction—the ability to discriminate between offenders who *will* and *will not* recidivate (emphasis mine)" (Northpointe Inc., 2015, p. 7). Northpointe's use of the indicative instead of the conditional tense confirms the imperceptible shift from probability to determinism previously discussed. This shift is enabled by overlooking the performative power of decisions. However, the connection established between prediction and determinism by the use of the future tense becomes more than an imprecise use of language when COMPAS gets used for the purpose of sentencing. By deciding, and thus performatively determining the present based on past data, one confirms the past state as the norm in light of which the future is preemptively understood as having taken place. Therefore, the past is that which will repeat itself simply because it once was the case.

In addition, the connection between a possible and an inescapable future is materially realized when, by modifying the future of an individual based on predictions, one creates new data that will eventually be added to the dataset serving to train the algorithm. As I am about to show, the more we judge on the basis of predictions, the more we produce auto-confirming data, and the more the reality will fit the data.

Let us draw on the case of the false positives: an individual is predicted to reoffend within 2 years but will not reoffend. This is nothing unexpected as individuals do defy the outcome predicted by the risk assessment tool in a high percentage of cases. In the case of COMPAS and of non-expert human beings' predictions, false positives for Black individuals amount to 40.4% as opposed to 25.4% for white individuals (Dressel & Farid, 2018, p. 2). I have already pointed out how social injustice and racism are reflected in these numbers. Let us now proceed with a thought experiment: the case of a male individual called Y. Y was excluded from probation measures because of his high risk to reoffend based on his history and on his COMPAS scores.[11] Instead, Y is sentenced to prison. What kind of data does this case produce? The data confirm the connection between the prediction of a high risk of recidivism and sentencing to prison time. But Y could have belonged to the class of false positives. The issue is that we will never know if this is the case, as Y was not granted probation, and was sent to prison. There is a good chance that Y is among the 40.4% of "outliers" who defy the predictions. However, the possibility of being an outlier to the prediction has been materially excluded by his prison sentencing. It is now impossible to find out if Y would have reoffended or not within 2 years, and thus impossible to rectify future predictions. Y's case cannot belong to the false positives anymore. As a result, decisions based on predictions systematically eliminate the false positive outliers. The predictions leading to a false negative are the only ones whose rectitude can be checked in real life. Let us pursue the thought experiment: male individual Z was predicted not to reoffend within 2 years and put on probation but ends up reoffending. Consequently, the result of the predictive algorithm was inaccurate and needs to be corrected, which in cybernetic terms is called "negative feedback."

[11] This has been the case in State of Wisconsin v. Eric L. Loomis, 2016, §19 for instance.

While the data produced in the case of Y solely confirms the correlation between prediction of recidivism and sentence to prison, the data produced in the case of Z will lead the algorithm to correct future predictions in order to avoid false negatives, leading to harsher predictions. This tendency will happen as soon as data resulting from algorithmic predictions are themselves integrated into the risk assessment tool. Following the mechanism of feedback loops, false positives are progressively eliminated, while false negatives lead to the correction of the algorithm, which results in more sentences to prison rather than releases on probation.

With the generalization of the use of predictive algorithms, no data will be produced that are not themselves the result of predictive mechanisms. The more data produced through predictive algorithms are fed back into the norm dataset, the more new predictions will reflect the absence of false positives and the necessity to avoid false negatives. As a result, less and less individuals should be predicted as non-reoffenders.

3.3 Justice as Risk Management

> By change what the policy deputies mean is contingency, risk, flexibility, and adaptability to the groundless ground of the hollow capitalist subject, in the realm of automatic subjection that is capital. […] This economy is powered by constant and automatic insistence upon the externalization of risk, the placement at an externally imposed risk of all life, so that work against risk can be harvested without end. Stefano Harney & Fred Moten, *The Undercommons: Fugitive Planning & Black Study,* pp. 76–77.

In my third and final argument, I aim to show that decision-making premised on predictive algorithms performs a certain understanding of the function of justice. In this understanding, justice is not about fairness; neither is it about retribution or rehabilitation. Rather, it functions as an apparatus for the biopolitical regulation of risks. The judged individual in their irreplaceable singularity is secondary to this purpose. The connection I would like to establish here is the one between biopolitics—understood as the management by the state of the life and death of a population—and a preemptive form of cybernetics. Traditionally, cybernetic systems are characterized by their self-regulation in order to maintain the stability of their organization against the tendency toward energetic dispersion or chaos called entropy (Wiener, 1989). The specificity of preemptive systems is that their regulation consists of anticipatively avoiding something that has not yet happened. In order to understand this mechanism, it is necessary to clarify what making a decision consists of, and what characterizes a decision based on a prediction.

The definition of decision I offer here is loosely based on Jacques Derrida's *Force of Law* (Derrida, 1990, p. 961f.) and Walter Benjamin's *Critique of Violence* (Benjamin, 1996). A decision etymologically consists in performing a cut (Latin, *decidere*) within a complex reality with the help of a calculation following a set of rules in order to determine what will or will not be. At the same time, a decision is composed of the interpretation of these rules and the results of the calculation based

on them. Indeed, if a decision were solely the result of a calculation, it would not be a decision. For instance, that 4 is the result of 2 + 2 is not a decision, only the result of a computation following a rule. To be a decision, a speech act must be more than computation.[12]

Let us unpack this attempt at a definition. While predictive algorithms like COMPAS compute the risk of an individual's recidivism, they contribute to but do not perform the decision strictly speaking as they provide a calculation without its interpretation. Predictions are expressed in the form of probabilities stretching from 0 (= will not happen) to 1 (= will happen). However, as demonstrated in the previous arguments, there cannot be a 0 or 1 probability of reoffending, as there is no way to gain absolute certainty regarding the future of a given individual. Because there can be no absolute certainty regarding the risk of an individual's recidivism, there can be no calculation of where to "make the cut." In contrast to predictions, the decision is binary. Deciding consists in turning the $x\%$ chance of an individual to reoffend into either a "will" (= 0) or "will not" (= 1) reoffend, and thus will or will not be sent to prison. The judge makes a cut by interpreting and evaluating the output of the algorithm.

Since there can be no certainty regarding the recidivism of an individual—no calculation of where to "make the cut"—the decision regarding the individual can never be entirely justified by computation and is, in that sense, ungroundable. Again, if a decision were to be fully grounded in computation, it would not be a decision (as in the case of 2 + 2 = 4). Therefore, each actual decision entails an interpretation with its measure of arbitrariness. It bears the risk of being wrong and comes with the responsibility associated with this risk. Because it can be wrong, a decision marks the limits of computation.

While predictive algorithms may mitigate the risk of making a wrong decision, they can never eliminate this risk because they cannot substitute for the measure of arbitrariness in every interpretation, which accounts for the never entirely groundable character of a decision. The use of predictive algorithms conceals that an interpretation happens each and every time a decision is made. This concealment contributes to the idea that justice could be reduced to the automatic application of rules.

We have previously established that a decision is performative. Let us now specify what is performed when a judicial decision is rendered. As we have seen, judicial decisions are performative in the sense that they reshape the life of the judged individual and the community that surrounds them. At the same time, in order to produce such an effect, the decision reaffirms the legal order and the context granting its performative power. A decision referring to an existing rule reinstates the legitimacy of the rule that serves to justify the decision—implying a circularity that underlines the necessary violence of the law (Derrida, 1990, p. 987; Benjamin, 1996). In consequence, deciding does not only entail determining what will be, it implies at the same time performatively reaffirming the

[12] As Derrida points out, a decision is free and thus responsible only if it is not solely the automatic application of a rule but interprets and thus "invents" the rule anew every time (Derrida, 1990, p. 961).

normative context in which the decision takes place and makes sense, a context without which the decision would have no legitimacy.

The normative context that is performatively reaffirmed by a decision based on predictive algorithms is risk management. By entrusting predictive algorithms to help make decisions in the judicial context, one displaces the idea of justice as that which is tied to an always-singular situation necessitating a specific interpretation of the law (Derrida, 1990, p. 948) toward, instead, an automatized mechanism of regulation and modulation of risks—be it the risks that criminality is considered to represent for society, or the risks tied to the consequences of a wrong or unfair decision. Ezekiel Dixon-Román et al. describe this management of risks in terms of cost minimization: "In other words, incorrectly identifying an individual as high-risk, and making decisions regarding the nature of that individual's probation and parole accordingly, is considered less costly than failing to identify someone who goes on to commit a 'serious offense' as defined above" (Dixon-Román et al., 2019, p. 31). How to explain this prioritization of risk management in judicial procedures over and above any regard for the judged individual? Antonia Majaca and Luciana Parisi conceive of the form of governmentality that makes use of predictive algorithms as "paranoid," tying it to the "white male subject of humanism" (Majaca & Parisi, 2016, p. 4). This kind of governmentality emerged from a sense of permanent threat tied to the 9/11 attack and is marked by the desire to act based on what is not known (Amoore, 2013, p. 55 f.). Lorraine Daston leads the generalization of predictive algorithms further back to the context of "Cold War rationality," where the risk of a nuclear catastrophe was mitigated by universal algorithmic procedures and the idea that everyone played the same game by the same rules. Daston describes "Cold War rationality" as a rationality relying on a set of rules that can be applied mechanically without interpretation, judgment, or deliberation.[13]

While this sense of paranoia and general suspicion inherited from the Cold War and 9/11 can partly explain the generalization of the use of predictive algorithms, I would argue that they are part of a biopolitical mechanism set in motion during the nineteenth century. Risk management—a logic nowadays shared by financial institutions, insurance companies, and the criminal justice system—functions as a "productive" tool (in the Foucauldian sense) for population management at the service of biopolitical governance.

The notion of risk management is connected to a cybernetic conception of society. Modulating risks is part and parcel of a society that, since the nineteenth

[13] Lorraine Daston, "The Rule of Rules, or How Reason Became Rationality," talk at the Wissenschaftskolleg zu Berlin, November 21, 2010. Referred to in Majaca and Parisi (2016). Appeared in the first chapter of the collective book (Erickson et al., 2013) under the title "Enlightenment Reason, Cold War Rationality, and the Rule of Rules.".

century, functions biopolitically. Here, we might remember that the aim of biopower is not to discipline bodies on the individual level; its goal is to establish regulating mechanisms from within the population in order to attain an equilibrium, "something like a homeostasis," writes Foucault, using cybernetic terminology (Foucault, 1997, p. 249).

From a cybernetic standpoint, living and mechanical processes obey the same logic: both are systems that regulate their relation to their environment through feedback mechanisms that enable them to maintain their internal organization against the system's tendency for energy dispersion or chaos (Wiener, 1989). Placed in this cybernetic context, predictive algorithms function as a naturalized means to maintain social homeostasis. The difference between traditional cybernetic systems and preemptive systems, however, is that in traditional cybernetic systems, the system regulates itself in light of events that have already happened and whose results are fed back into the system in order for it to adapt to a changing situation. By anticipating risks, preemptive systems regulate themselves relative to that which has not happened yet. They exclude in advance any event that could imperil an already given equilibrium, or more precisely, the norm that is at work in this equilibrium. And in order to protect themselves from hypothetical future risks, preemptive systems agree to expose the disenfranchised to actual risks in the present—be it the risks tied to predictive policing (Harcourt, 2007), to a life in prison, to unpayable health insurance, to homelessness and poverty.

As cited in the epigraph of this section, Stefano Harney and Fred Moten emphasize in *The Undercommons* that neoliberal capitalism is a mode of governance which submits disenfranchised, precarious Black and Brown lives to increasingly higher levels of contingency and flexibility—putting these lives at risk and making any kind of autonomous organization and planning increasingly difficult.[14] Similarly, in *Society Must Be Defended* (Foucault, 2003), Foucault describes racism as the way for biopower to let a part of the unwanted population die by exposing it to multiple risks of death or to political death by exclusion. By the sustained exposure of disenfranchised populations to risks by means of risk management tools, the government exerts its biopolitical prerogative to let die in order to maintain its perverse homeostasis.

4 Conclusion: for Abolition

Races are not a given. Instead, even construed as biological as is still often the case in the USA (Vyas et al., 2020), they are the products and effects of biopolitical technologies of differentiation and hierarchization

[14] See also Martin (2011, p. 260).

applied to populations. Predictive algorithms are not only iteratively performing racial biases; they are producing the conditions of possibility for racial differences and hierarchies by automatically maintaining the disenfranchised in the never-ending present of crime, prison, and political exclusion, while the open future remains reserved for the privileged population still catered to on an individual basis. Determinism, here, is not the expression of the accuracy of the knowledge that complex algorithms would have gained over individuals—much like the series *Minority Report* and *Westworld* would present it. Rather, understood as the denial of an individual's right to a degree of self-determination, determinism is manufactured through predictive decision-making. By algorithmically excluding futures based on past data, predictive tools are at their very core conservative. Predictive algorithms are technological means of production and reproduction of social homeostasis, transforming society into a program, an algorithm that flattens the future of the disenfranchised into one inescapable present. In consequence, the abolition of predictive algorithms is necessary.

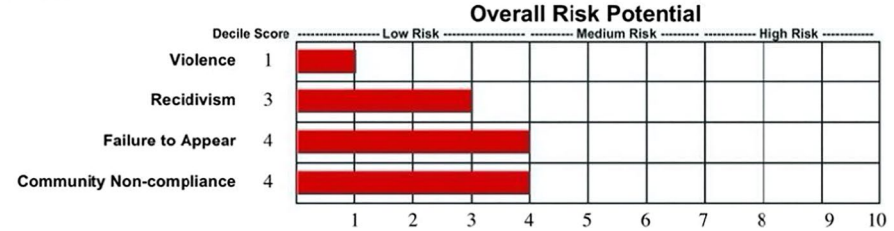
As argued from the beginning of this paper, abolishing a dispositive never solely means repealing the concrete material stuff from which the dispositive is made—here, the risk assessment tools—but in addition, repealing the social conditions that give the dispositive its apparent necessity. At issue in the present case is a society that accepts to systematically expose the disenfranchised to actual risks in the present in order to avoid hypothetical future risks in order to maintain its equilibrium. Abolishing the "program" of current society, which obeys a cybernetic-preemptive logic, is the only way forward, toward the possibility of opening futures in the now. Abolishing is a call to imagine justice otherwise.

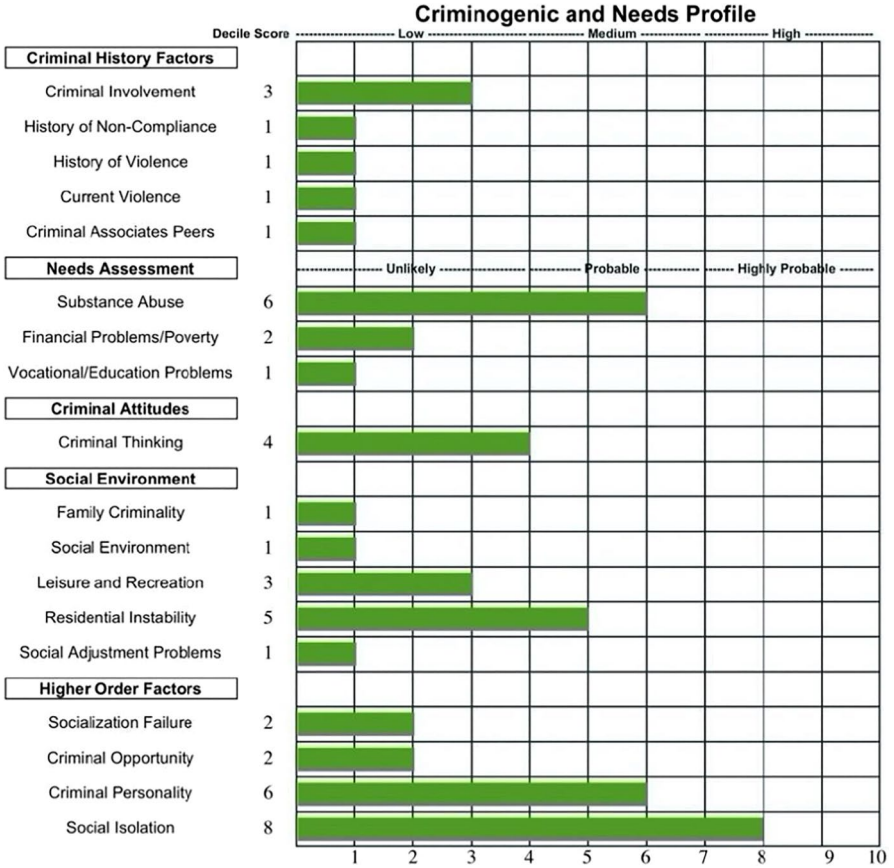
Appendix

Northpointe COMPAS risk assessment, offender #: 01cr57 (August 14, 2006). Retrieved on November 20, 2018, from https://assets.documentcloud.org/documents/2839240/Sample-Risk-Assessment-COMPAS-Results.pdf

Northpointe COMPAS Risk Assessment

Name: **Class3, Jessie** SSN: Offender #: **01cr57**
Date of Birth: **06/19/1977** Date of Screening: **08/14/2006**
Comment:





Acknowledgements I would like to thank Deanna Cachoia-Schanz for her help, reading, commenting, and editing this paper over the long period of time it took me to write it. Thank you to Christine Allen-Blanchette and Jeremy Gallion for our discussions on the topic of predictive algorithms back in Philadelphia in 2018. Thank you to Prof. Lisa Miracchi and the members of the MIRA group as well as to the Theorizing Colloquium (University of Pennsylvania) for allowing me to present and discuss my paper. Finally, thank you to Kristen Tapson for the finishing touch.

I want to thank both reviewers for their time and their valuable comments. This paper has become stronger as a result.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Amoore, L. (2013). *The politics of possibility: Risk and security beyond probability*. Duke University Press.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias. There's software used across the country to predict future criminals. And it's biased against blacks*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Austin, J. L. (1962). *How to do things with words: The William James lectures*. Clarendon Press.

Barabas, C., Dinakar, K., Ito, J., Virza, M., & Zittrain, J. (2017). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. *arXiv*, 1712.08238v2. http://arxiv.org/abs/1712.08238v2

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Bateson, G. (1979). *Mind and nature: A necessary unity*. E. P. Dutton.

Beckett, K. (1999). *Making crime pay. Law and order in contemporary American politics*. Oxford University Press.

Benjamin, R. (2019). *Race after technology*. Polity Press.

Benjamin, W. (1996). Critique of violence. In M. W. Jennings (Ed.), *Selected writings volume 1. 1913–1926* (pp. 236–252). Harvard University Press. https://doi.org/10.1515/9780822390169-037/html

Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale J. l. & Tech., 20*, 103–176.

Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior, 36*(1), 21–40. https://doi.org/10.1177/0093854808326545

Campolo, A., & Crawford, K. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, *6*, 1. https://doi.org/10.17351/ests2020.277

Card, D., & Smith, N. A. (2020). On consequentialism and fairness. *Frontier in Artificial Intelligence, 3*(34), 1–11. https://doi.org/10.3389/frai.2020.00034

Chaitin, G. (2005). *Meta math!* Vintage Books.

Clear, T. (2008). The effects of high imprisonment rates on communities. *Crime and Justice, 37*(1), 97–132. https://doi.org/10.1086/522360

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). *Algorithmic decision making and the cost of fairness*. In. New York, NY, USA: ACM. http://dx.doi.org/https://doi.org/10.1145/3097983.3098095

Couzens, R. J. (2011). *Evidence-based practices. Reducing recidivism to increase public safety: A cooperative effort by courts and probation*.

Davis, A. Y. (2005). *Abolition democracy: Beyond prisons, torture, empire*. Seven Stories Press.

Derrida, J. (1990). Force of law. The "mystical foundation of authority". *Cardozo Law Review*, *11*(5/6), 920–1045.

Dixon-Román, E., Nyame-Mensah, A., & Russell, A. R. (2019). Algorithmic legal reasoning as racializing assemblages. *Computational Culture*, *7*. http://computationalculture.net/algorithmic-legal-reasoning-as-racializing-assemblages/

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1), 1–5. https://doi.org/10.1126/sciadv.aao5580

Erickson, P., Klein, J. L., Daston, L., Lemov, R., Sturm, T., & Gordin, M. D. (2013). *How reason almost lost its mind: The strange career of Cold War rationality*. The University of Chicago Press.

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

Foucault, M. (1997). *« Il faut défendre la société ». Cours au Collège de France (1975–1976)*. Seuil, Gallimard.

Foucault, M. (2003). *"Society must be defended": Lectures at the Collège de France, 1975–1976* (D. Macey, Trans.). Picador.

Foucault, M. (2012). *Discipline & punish: The birth of the prison* (A. Sheridan, Trans.). Vintage Books.

Hamilton, M. (2014). Risk-needs assessment: Constitutional and ethical challenges. *American Criminal Law Review, 52*(231), 231–291.

Harcourt, B. E. (2007). *Against prediction: Profiling, policing, and punishing in an actuarial age*. The University of Chicago Press.

Harney, S., & Moten, F. (2013). *The undercommons: Fugitive planning and black study*. Minor Compositions.

Hayles, N. K. (2012). *How we think: Digital media and contemporary technogenesis*. The University of Chicago Press.

Hoffman, P. B., & Adelberg, S. (1980). The salient factor score: A nontechnical overview. *Federal Probation, 44*, 44–52.

Holland, T. E. (1881). *The institutes of Justinian*. Clarendon Press.

Klingele, C. (2016). The promises and perils of evidence-based corrections. *Notre Dame Law Review, 91*(2), 537–584.

Lin, Z. J., Jung, J., Goel, S., & Skeem, J. (2020). The limits of human predictions of recidivism. *Science Advances, 6*(7), 1–8. https://doi.org/10.1126/sciadv.aaz0652

Majaca, A., & Parisi, L. (2016). The incomputable and instrumental possibility. *e-flux journal*, *77*.

Martin, R. (2011). From the race war to the war on terror. In P. Clough Ticineto & C. Willse (Eds.), *Beyond biopolitics: Essays on the governance of life and death* (pp. 258–274). Duke University Press.

Mbadiwe, T. (2018). Algorithmic injustice. *The*. *New Atlantis, 54*, 3–28. https://doi.org/10.2307/90021005

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Northpointe Inc. (2015). *Practitioner's guide to COMPAS core*. Northpointe Inc.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Siegert, B. (2013). Cultural techniques: Or the end of the intellectual postwar era in German media theory. *Theory, Culture & Society, 30*(6), 48–65. https://doi.org/10.1177/0263276413488963

State of Wisconsin v. Eric L. Loomis. *2015AP157-CR, 2016 WI 68*.

Tonry, M. (2014). Legal and ethical issues in the prediction of recidivism. *Federal Sentencing Reporter, 26*(3), 167–176.

Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—Reconsidering the use of race correction in clinical algorithms. *The New England Journal of Medicine, 383*(9), 874–882. https://doi.org/10.1056/NEJMms2004740

Wiener, N. (1989). *The human use of human beings*. Cybernetics and society.

Wong, P.-H. (2020). Democratizing Algorithmic Fairness. *Philosophy & Technology, 33*(2), 225–244. https://doi.org/10.1007/s13347-019-00355-w