in LINKEDIN

Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing

By Andrew Lee Park / In Law Meets World / February 19, 2019 / 21 Min read / Add comment

Introduction

The notion of crime-prediction technology has been explored in science fiction for quite some time. Though first rearing its head in Phillip K. Dick's 1956 short story, "The Minority Report," it remained somewhat dormant in popular consciousness until a half-century later, when Steven Spielberg brought the story to the screen as a thrilling tech-noir blockbuster. The premise was simple: In the year 2054, crime in America is all but nonexistent thanks to the advent of "PreCrime" technology, which allows law enforcement to "see" a crime before it occurs.

In the opening scene, Chief John Anderton of the PreCrime Division, played by the inimitable Tom Cruise, is presented with an urgent PreCrime report—that later that day, a man was to violently stab his wife to death in their home. Immediately, the authorities mobilize like clockwork, and within minutes we see Anderton burst through the front door of the home. There, he finds a dowdy man of middle age standing in the living room, a large pair of scissors in his hands. Seated before him are two lovers—his unfaithful wife and her paramour, caught in the act. Our hero rushes to accost the man in just the nick of time, scarcely a moment before the predicted killing would be realized. As the man is whisked away by authorities for the



ABOUT

SCHOLARSHIP SUBMISSIONS EVENTS

FACEBOOK

LINKEDIN

Yet the scene evokes a shade of dystopian anxiety, even assuming that the murder was certain to occur. Being punished and condemned for an unrealized crime offends our ideas of blameworthiness, of moral agency and free will. But lucky for us, we would strain to imagine an America that would allow for such a system to exist. Constitutional protections, as well as the limits of technology as far as we know, would presumably bar PreCrime-like measures from ever being enacted.

But if we were to imagine, for just a moment, that we are indeed living in such a timeline, we might also imagine that the seeds of this future were planted in the summer of 2016. It was then that State v. Loomis was decided, in which the Wisconsin Supreme Court upheld the use of a particular algorithm in judicial decisionmaking—a form of machine intelligence that had become an indispensable part of Wisconsin's sentencing procedure. The algorithm at issue, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), was designed to assess a defendant's risk of recidivism—that is, the potential risk that the defendant will commit a crime in the future.²

I. The Use of COMPAS in Sentencing

In the 1990s, a company called Northpointe, Inc. set out to create what is now known as COMPAS, a statistically-based algorithm designed to assess the risk that a given defendant will commit a crime after release. In 2012, after years of development, Wisconsin implemented COMPAS into its state sentencing procedures, at which point COMPAS assessments officially became a part of a defendant's presentence investigation (PSI) report.

COMPAS's algorithm uses a variety of factors, including a defendant's own responses to a lengthy questionnaire, to generate a recidivism-risk score between 1 and 10.5 In gen



ABOUT SCHOLARSHIP

SUBMISSIONS EVENTS

FACEBOOK

LINKEDIN

least some degree—by COMPAS's recidivism risk assessment.

But in more precise terms, how does COMPAS calculate its risk score? And what specific kinds of data does it consider? Surprisingly, aside from Northpointe, no one truly knows. One would expect, at minimum, that the court implementing COMPAS would understand how it functions, but that is sadly not the case. In a concurring opinion in Loomis, Justice Shirley S. Abrahamson bemoans this point, noting that "the court repeatedly questioned both the State's and defendant's counsel about how COMPAS works. Few answers were available."8 Though Northpointe has been asked to reveal COMPAS's source code, it has staunchly refused to do so.⁹ And as troubling as it sounds, such a refusal is squarely within Northpointe's legal rights.

As a privately developed algorithm, COMPAS is afforded the protections of trade secret law. $\frac{10}{100}$ That means that COMPAS's algorithm—including its software, the types of data it uses, and how COMPAS weighs each data point—is all but immune from third-party scrutiny. 11 This extends not only to those who might exploit the algorithm for pecuniary gain. This also applies to the prosecutors who are putting forth sentencing recommendations, to the defendants who are sentenced under consideration of COMPAS scores, and—as Kafka rolls furiously in his grave—to the judges who use those scores in their sentencing decisions. $\frac{12}{12}$

Put simply, COMPAS answers to no one but its creators. Perhaps if COMPAS could at least be demonstrated to treat all defendants fairly and reliably, there may be less cause for concern. But as a recent study has shown, that is almost certainly not the case.

II. Bias in the Machine

A. Competing Notions of Fairness



SCHOLARSHIP SUBMISSIONS EVENTS ABOUT

FACEBOOK

more likely than white defendants to be incorrectly judged at a higher risk of recidivism, and (2) white defendants were more likely than black defendants to be incorrectly judged as low risk. 14 The study was criticized by Northpointe in a 37-page defense, which in turn was rebutted by ProPublica soon after. 15 Northpointe contended that its scores were fair because its rate of accuracy in predicting recidivism was the same for black and white defendants—about 60 percent. ¹⁶ This contention is true, and even ProPublica does not deny that figure. ¹⁷ But ProPublica also stuck by its findings that COMPAS was unfair in treating black and white defendants differently in terms of its *incorrect* scores. So how can a score be both fair and unfair at the same time?

A group of researchers studied this phenomenon and published their findings in a Washington Post blog. 18 The problem is not explicitly about race, as COMPAS purportedly does not use race as a factor in its risk score. 19 The problem, according to the researchers, is about competing notions of what constitutes fairness in the first place. COMPAS's model attempts to capture fairness in terms of accurately predicting those who do, in fact, reoffend; as an example, those who received a score of seven went on to reoffend roughly 60 percent of the time, regardless of race. However, ProPublica moves the focus away from reoffenders and toward those who end up not reoffending. Under this lens, blacks were roughly twice as likely as whites to be mistakenly classified as medium or high risk, even though they ultimately end up on the straightand-narrow after release.

But here is the problem: It is mathematically impossible for a model to satisfy both fairness criteria at the same time. Correcting for one necessarily erodes the accuracy of the other. 20 As long as COMPAS calibrates its algorithm according to its notion of fairness, the incongruity noted by ProPublica will inevitably occur. This leads us to an important question: Are the benefits of COMPAS getting it right worth the costs to black defendants when it gets it wrong? It is clear how



SCHOLARSHIP SUBMISSIONS ABOUT

EVENTS

FACEBOOK

B. Structural blases in the Data Itself in

The worry doesn't stop there, unfortunately. As stated above, race is purportedly not a factor in COMPAS's algorithm. But there are likely other factors COMPAS analyzes that serve as *proxies* for race, which may lead to racial bias in results. Even seemingly innocuous data points can exert prejudice against marginalized demographics. Consider, for example, one's area of residence. Heavier policing in minority-dominated often inflate arrest statistics for individuals residing in those areas.²¹ As COMPAS is a statistics-based algorithm, this could make black defendants more prone to inordinately high-risk scores by virtue of where they live.

Such biases would not necessarily be caused by a racist algorithm, as algorithms can largely be characterized as number crunchers. The problem is that algorithms are trained on data produced by humans.²² If the data itself is tainted with historic and structural biases, that taint will necessarily be imputed onto an algorithm's output. Algorithms in other contexts have already demonstrated this phenomenon, with gender stereotyping skewing Google Image searches, 23 racial stereotyping influencing the appearance of targeted advertisements, 24 and gay stereotyping leading to absurd recommendations in Google Play's algorithm. 25

As a result, there is a growing fear that algorithms can cause "runaway feedback loops." This is where historic biases are reflected in an algorithm's result, further skewing data against marginalized groups, which is then processed again by algorithms to produce even more biased results. 26 Viewing COMPAS through this lens, the situation grows dire. The immediate risk is that overrepresented minorities will be issued errant risk scores at an ever-increasing rate. Looking beyond that, however, we can see how longer sentences for certain groups can push them further into poverty and joblessness, into deeper marginalization and disillusionment.



ABOUT SCHOLARSHIP SUBMISSIONS EVENTS

FACEBOOK

Northpointe's defense was to contend that COMPAS was accurate under its own measure of fairness. To lean on such a defense, unfortunately, is to burn ants with a looking glass in one hand while pointing to its shadow with the other.

III. Loomis's Blind Spot: Heuristics and Cognitive **Biases**

A. How the *Loomis* Court Got It Wrong

If COMPAS does indeed discriminate based on race, should we be afraid that sentencing decisions will carry the taint of racial prejudice? The Loomis court certainly didn't think so. The court acknowledged ProPublica's study in one breath and discounted it in another. ²⁸ And though the court failed to properly assess COMPAS's reliability, the court's more egregious oversight was in failing to consider the fallibility of the relevant decisionmakers at issue: the sentencing judges.

Of all the arguments raised by the defendant in *Loomis*, there are two due process challenges that are particularly important: first, that the use of a COMPAS risk assessment at sentencing violates the right to be sentenced based upon accurate information, and second, that it violates a defendant's right to an individualized sentence. 29 These challenges were grounded in two fundamental aspects of COMPAS assessments: that trade secret protections prevent defendants from verifying a risk score's accuracy, and that COMPAS uses group data to determine its score.³⁰

In responding to the first challenge, the court misses the mark by a mile. The court held that despite COMPAS's trade secret protections, a defendant could verify that she correct



ABOUT SCHOLARSHIP SUBMISSIONS

EVENTS

FACEBOOK

LINKEDIN

But more significant is the court's response to the second challenge. It held that the use of COMPAS did not violate the defendant's right to an individualized sentence based on three assumptions: (1) The COMPAS score, despite its use of group data, was not the determinative factor at sentencing; (2) risk scores give judges more "complete information" which allows them to better weigh sentencing factors; and (3) trial courts can "exercise discretion" when assessing a defendant's score, and as a result, disregard scores that are inconsistent with a defendant's other factors. 32 How the *Loomis* court arrived at these assumptions is unclear, to say the least. Recall that sentencing judges do not have access to COMPAS's algorithm. Like everyone else, they only have access to the naked score, which is appended to the back of their PSI report. 33 The court assumes that judges—without any qualifications regarding COMPAS—will not overrely on a score, will weigh the other sentencing factors fairly, and will know when a score is wrong. But if the wealth of studies regarding cognitive biases is any indication, these assumptions are misguided.

B. COMPAS and Cognitive Bias

Research has established that even the most prudent decisionmakers are prone to severe errors in judgment. These cognitive biases result from the brain's natural tendency to rely on heuristics, or simple rules of thumb, when dealing with complicated mental tasks. 34 Though there are several types of cognitive biases that can cause a judge to over-rely on a COMPAS score, there is one that is particularly salient, which is known as *automation bias*.

Automation bias refers to the tendency to "ascribe greater power and authority to automated aids than to other sources of advice." Studies show that automation bias rears its head in a wide variety of situations, from evacuees blindly following malfunctioning robots in emergency situations, 36 to seasoned radiologists relying on faulty diagnostic aids when they wol Type here to search...



SCHOLARSHIP ABOUT

SUBMISSIONS

EVENTS

FACEBOOK

LINKEDIN

So was the *Loomis* court right in believing COMPAS scores would not become a determinative factor at sentencing? Or that judges would know when to disregard an errant score? Almost certainly not. Given the inherent complexities and time constraints of sentencing, judges are prone to place undue weight on a COMPAS score. This is exacerbated by the fact that COMPAS's manual informs judges that a "counter-intuitive risk assessment" is not an indicator that the algorithm has functioned improperly. 40

Other cognitive biases serve to further entrench COMPAS's role in sentencing decisions. Confirmation bias, or the tendency to seek information that validates one preconceived notion while rejecting other information, ⁴¹ may lead judges to disregard other factors in a PSI report that counter a given risk score. Bias blind spot, or the tendency for people to see themselves as less susceptible to bias than others, 42 may cause judges to underestimate the degree to which their sentencing decisions are being skewed by COMPAS assessments. As a result, it is not only likely that defendants' due process rights are violated at sentencing when COMPAS is involved, but it is also likely that sentencing judges may be acting, inadvertently, as facilitators of racial prejudice as a result of this overreliance.

Conclusion

In the final act of *Minority Report*, PreCrime is abolished after the discovery of a crucial flaw once potential criminals become aware of their future, they have the power to avert it. In our world, however, defendants assessed under COMPAS remain at the mercy of its imperfections. 43 The ramifications of this are clear: Crime-prediction technology like COMPAS will continue to disfavor marginalized groups, judges will continue reinforcing this bias through their sentencing



ABOUT

SCHOLARSHIP SUBMISSIONS EVENTS

FACEBOOK

Seeing now prevalent at has become in our everyday lives, predictive algorithms like contras will likely become increasingly common in our criminal justice system. However, the outlook is not all doom and gloom. When wielded correctly, Al may indeed promote efficiency without implicating significant concerns related to machine bias. Thus, the solution must be more nuanced than to banish them entirely.

For one, our laws must adapt to the novel challenges these technologies present. Specifically, trade secret law should not serve to bar defendants from raising and investigating valid due process questions. As scholars have noted, the law and policy governing trade secrets must be reformed to account for the individual rights and social interests at stake. 45 Protections meant to safeguard a company's economic interests should not be blindly applied where legitimate issues of social justice are implicated.46

In addition, data scientists must work with the state to develop ways to mitigate algorithmic prejudice and potential feedback cycles. It would be difficult to fully strip data of its structural and historical biases. However, research can be done to flag the types of data points that are most prone to racial bias. Algorithms can then be trained to weigh those data points less heavily, or to compute them in a manner that minimizes the probability of promoting racial disparity. 47

But most of all, there must be a broader discussion among policymakers regarding the role machine intelligence should play in our judicial system as a whole. Fortunately, this discussion need not start from scratch. At a developer conference in 2016, Microsoft CEO Satya Nadella shared his approach to AI. First, machine intelligence should augment rather than displace human decisionmaking; second, trust must be built directly into new technology by infusing them with modes of transparency and accountability; and third, technology must be inclusive and respectful to everyone. 48 Principles such as these must anchor and guide future policy



ABOUT SCHOLARSHIP SUBMISSIONS EVENTS

FACEBOOK

As there are signs we are eugling towards a Prechine-like future in Tiple ways than one, we find ourselves at a critical juncture. If we proceed without due consideration of the risks these algorithms pose, we may find ourselves relying far too much on technologies we do not fully understand. We may unwittingly begin perpetuating past injustices on a widespread, systematic level. And chillingly, we may be headed for a future where individuals are regularly condemned for prospective crimes they may never commit. And that, we may agree, is a future best left for the realm of science fiction.

² Jennifer L. Skeem & Jennifer Eno Louden, Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) 8 (2007), http://risk-resilience.berkeley.edu/sites/default/files/journalarticles/files/assessment_of_evidence_on_the_quality_of_the_correctional_offender_management_profiling_for_alternative_sanctions_ Northpointe, Inc., Practitioner's Guide to COMPAS Core 1 (2015), http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf.

³ Id. Northpointe, Inc. recently changed its name to "Equivant" after a 2017 rebranding effort. Courtview, Constellation & Northpointe Re-Brand to Equivant, Equivant, http://www.equivant.com/blog/we-have-rebranded-to-equivant. For the sake of convenience, this piece continues to refer to the company as Northpointe.

https://www.wisbar.org/NewsPublications/InsideTrack/Pages/Article.aspx? Volume=9&Issue=14&ArticleID=25730#3.

¹ 881 N.W.2d 749 (Wis. 2016).

⁴ Joe Forward, *The* Loomis *Case: The Use of Propriety Algorithms at Sentencing*, State Bar Wis.: Inside Track (July 19, 2017),



SCHOLARSHIP ABOUT

SUBMISSIONS

FACEBOOK

EVENTS ⁶ Northpointe, Inc., *supra* note 2, at 31. ⁷ *Id.* at 8. ⁸ State v. Loomis, 881 N.W.2d 749, 774 (Wis. 2016) (Abrahamson, J., concurring). ⁹ Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice* System, 70 Stan. L. Rev. 1343, 1346 (2018).

- ¹⁰ *Loomis*, 881 N.W.2d at 761.
- ¹¹ See Taylor R. Moore, Ctr. for Dem. & Tech., Trade Secrets and Algorithms as Barriers to Social Justice (2017), https://cdt.org/files/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf.
- ¹² See Loomis, 881 N.W.2d at 774 (Abrahamson, J., concurring) (noting that neither State nor defendant's counsel could explain to a Loomis court judge how COMPAS worked).
- ¹³ See Julia Angwin et al., Machine Bias, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
- ¹⁴ Jeff Larson et al., How We Analyzed the COMPAS Recidivism Algorithm, ProPublica (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.



SCHOLARSHIP ABOUT

SUBMISSIONS EVENTS

f FACEBOOK

julia Aligwili ∝ jeli Larsoli, *Dias III Crillillal Scores is Wathelliaucally Illeytable, kesearchers* In Linkedin Say, ProPublica (Dec. 30, 2016, 4:44 PM), https://www.propublica.org/article/bias-in-criminal-riskscores-is-mathematically-inevitable-researchers-say.

¹⁷ *Id.*

¹⁸ Sam Corbett-Davies et al., A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear., Wash. Post: Monkey Cage (Oct. 17, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithmbe-racist-our-analysis-is-more-cautious-than-propublicas/? noredirect=on&utm term=.fa2ecb6503a8.

¹⁹ *Id.*

²⁰ *Id.*

²¹ *Id.*

²² Jeremy Kun, *Big Data Algorithms Can Discriminate, and It's Not Clear What to Do About It,* Conversation (Aug. 13, 2015, 1:56 AM), https://theconversation.com/big-data-algorithms-candiscriminate-and-its-not-clear-what-to-do-about-it-45849.

²³ See Jennifer Langston, Who's a CEO? Google Image Results Can Shift Gender Biases, U. Wash. News (Apr. 9, 2015), https://www.washington.edu/news/2015/04/09/whos-a-ceo-google-imageresults-can-shift-gender-biases/.



ABOUT SCHOLARSHIP SUBMISSIONS EVENTS

f FACEBOOK

[25] See Wilke Arianny, The Curious Connection between Apps for Gay Ivieti and Sex Offenders, Atlantic (Apr. 14, 2011), https://www.theatlantic.com/technology/archive/2011/04/the-curiousconnection-between-apps-for-gay-men-and-sex-offenders/237340/.

[26] See Danielle Ensign et al., Runaway Feedback Loops in Predictive Policing, 81 Proc. Machine Learning Res. 1 (2018).

[27] See Angwin et al., supra note 13 (noting that Tim Brennan, Northpointe's founder, admitted that creating a risk score that did not "include items that can be correlated with race—such as poverty, joblessness and social marginalization"—would be difficult).

[28] State v. Loomis, 881 N.W.2d 749, 761 (Wis. 2016).)

[29] *Id.* at 761.

[30] *Id.* at 761–65.

[31] *Id.* at 761.

[32] *Id.* at 764–65.

[33] See, e.g., Wisc. Dep't of Corr., Compas PSI Presentation, http://www.wispd.org/attachments/article/272/COMPAS%20PSI%20Presentation%20by%20DOC.pdf.

[34] Amos Tversky & Daniel Kahneman, Judgment Under Uncertainty: Heuristics and Biases, 185 Science 1124, 1125 (1974). Type here to search...



ABOUT SCHOLARSHIP

SUBMISSIONS EVENTS

FACEBOOK

[30] See Paul Rodinielle et al., Overtrust of Rodols III Efficigency Evacuation Scenarios, 11th ACM/IEEE Int'l Conf. on Hum-Robot Interaction 101 (2016).

[37] Parasuraman & Manzey, supra note 35, at 394.

[38] *Id.* at 392.

[39] Kate Goddard et al., Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators, 19 J. Am. Med. Informatics Ass'n 121, 125 (2012).

[40] Northpointe, Inc., supra note 2, at 28.

[41] Stephen Porter & Leanne ten Brinke, Dangerous Decisions: A Theoretical Framework for Understanding How Judges Assess Credibility in the Courtroom, 14 Legal & Crim. Psych. 119, 127 (2009).

[42] Emily Pronin & Matthew B. Kugler, Valuing Thoughts, Ignoring Behavior: The Introspection Illusion as a Source of the Bias Blind Spot, 43 J. Experimental Soc. Psych. 565, 565 (2007).

[43] See Jason Tashea, Risk-Assessment Algorithms Challenged in Bail, Sentencing, and Parole Decisions, ABA J. (Mar. 2017),

http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole (noting that COMPAS continues to be used in several jurisdictions).

[44] See Christopher Rigano, Using Artificial Intelligence to Address Criminal Justice Needs, Nat'l

Inst. Just. (Oct. 8, 2018), https://www.nij.gov/journals/280/Pages/using-artificial-intelli



ABOUT

SCHOLARSHIP SUBMISSIONS EVENTS

FACEBOOK

[45] See, e.g., Moore, Supra note 11 (urging that intellectual property law supra note 11) that "there is an equitable balance between people's liberty interests . . . and a company's interest in maintaining its trade secret").

[46] See id. ("[W]hile [trade secret law] allows companies to securely deploy cutting-edge software in various fields, it can simultaneously perpetuate and exacerbate existing discriminatory social structures when these systems go unchecked and unregulated.").

[47] Chris DeBrusk, The Risk of Machine-Learning Bias (and How to Prevent It), MIT Sloan Mgmt. Rev. (Mar. 26, 2018), https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-andhow-to-prevent-it/.

[48] Satya Nadella, The Partnership of the Future, Slate (June 28, 2016, 2:00 PM), https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-worktogether-to-solve-societys-challenges.html.

[49] In addition to the use of algorithms in the courts, law enforcement agencies have begun rolling out privately developed crime-prediction algorithms that forecast when and where crime is likely to occur. Mark Smith, Can We Predict When and Where a Crime Will Take Place?, BBC (Oct. 30, 2018), https://www.bbc.com/news/business-46017239. The algorithm uses geographical data and crime statistics to inform officers where they should spend their time patrolling. Id.

Share this:

in LinkedIn Print **f** Facebook Twitter



ABOUT SCHOLARSHIP

SUBMISSIONS

EVENTS

FACEBOOK

n LINKEDIN

READ MORE

Blockchain-based Digital Assets and the Case for Revisiting Copyright's First Sale Doctrine

Various concerns about the reproducibility of intangible items have kept digital files from receiving the protections provided by the first sale doctrine of the Copyright Act of 1976. This piece breaks down blockchain technology and discusses how...

By Phillip Shaverdian / In Law Meets World / February 19, 2019

Emerging Digital Technology and the "Law of the Horse"

Professor Solow-Niederman identifies the themes that run through the series, explores the challenges that disruptive technologies pose to current law and policy, and points to the human values at stake with each development and intervention.

By Professor Alicia Solow-Niederman / In Law Meets World / February 19, 2019

Blockchain Technology and the Government: Dealing With the Threat of Data Manipulation and Increasing Records Longevity

Issues with record-keeping and management of records has long been a barrier to efficient administration within the government. This piece provides an overview of the technology behind blockchain and explores how it can be used to facilitate record...

Destination Unknown: The Perilous Future of Blockchain and Artificial Intelligence Technologies Under the California Consumer Privacy Act of 2018

This piece elucidates how the California Consumer Privacy Act of 2018 may serve to stifle widespread commercial applications of technologies

utilizing Blockchain and artificial intelligence. Though the requirements under the Act are designed to...

SEARCH



ABOUT SCHOLARSHIP

SUBMISSIONS

EVENTS

f FACEBOOK

in LINKEDIN

Copyright © 2019 UCLA Law Review - All Rights

Reserved

Website Supported by Archipelago Web