AN UNEASY DANCE WITH DATA: RACIAL BIAS IN CRIMINAL LAW

JOSEPH J. AVERY*

INTRODUCTION

Businesses and organizations expect their managers to use data science to improve and even optimize decisionmaking. The founder of the largest hedge fund in the world has argued that nearly everything important going on in an organization should be captured as data.[1] Similar beliefs have permeated medicine. A team of researchers has taken over 100 million data points from more than 1.3 million pediatric cases and trained a machine-learning model that performs nearly as well as experienced pediatricians at diagnosing common childhood diseases.[2]

Yet when it comes to some criminal justice institutions, such as prosecutors' offices, there is an aversion to applying cognitive computing to high-stakes decisions. This aversion reflects extra-institutional forces, as activists and scholars are militating against the use of predictive analytics in criminal justice.[3] The aversion also reflects prosecutors' unease with the practice, as many prefer that decisional weight be placed on attorneys' experience and intuition, even though experience and intuition have contributed to more than a century of criminal justice disparities.

Instead of viewing historical data and data-hungry academic researchers as liabilities, prosecutors and scholars should treat them as assets in the struggle to achieve outcome fairness. Cutting-edge research on fairness in machine learning is being conducted by computer scientists,

* Joseph J. Avery is a National Defense Science & Engineering Graduate Fellow at Princeton University; Columbia Law School, J.D.; Princeton University, M.A.; New York University, B.A.

1. RAY DALIO, PRINCIPLES: LIFE AND WORK 527 (2017).

2. Huiying Liang et al., *Evaluation and Accurate Diagnoses of Pediatric Diseases Using Artificial Intelligence*, 25 NATURE MED. 433, 433 (2019), https://www.nature.com/articles/s41591-018-0335-9.pdf.

3. Karen Hao, *AI is Sending People to Jail—and Getting It Wrong*, MIT TECH. REV. (Jan. 21, 2019), https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai.

applied mathematicians, and social scientists, and this research forms a foundation for the most promising path towards racial equality in criminal justice: suggestive modeling that creates baselines to guide prosecutorial decisionmaking.

I. PROSECUTORS AND RACIAL BIAS

More than 2 million people are incarcerated in the United States, and a disproportionate number of these individuals are African American.[4] Most defendants—approximately 95%—have their cases resolved through plea bargaining.[5] Prosecutors exert tremendous power over the plea bargaining process, as they can drop a case, oppose bail or recommend a certain level of bail, add or remove charges and counts, offer and negotiate plea bargains, and recommend sentences.[6]

When it comes to racial disparity in incarceration rates, much of it can be traced to prosecutorial discretion. Research has found that prosecutors are less likely to offer black defendants a plea bargain, less likely to reduce their charge offers, and more likely to offer them plea bargains that include prison time.[7] Defendants who are black, young, and male fare especially poorly.[8]

One possible reason for suboptimal prosecutorial decisionmaking is a lack of clear baselines. In estimating the final disposition of a case, prosecutors have very little on which to base their estimations. New cases are perpetually commenced, and prosecutors must process these cases quickly and efficiently, all while receiving subpar information; determining

4. Danielle Kaeble & Mary Cowhig, *Correctional Populations in the United States, 2016*, BUREAU JUST. STAT. 1 (Apr. 2018), https://www.bjs.gov/content/pub/pdf/cpus16.pdf.

5. Lindsey Devers, *Plea and Charge Bargaining: Research Summary*, BUREAU JUST. ASSISTANCE 1 (Jan. 24, 2011), https://www.bja.gov/Publications/PleaBargainingResearchSummary.pdf. Plea bargaining is a process wherein a defendant receives less than the maximum charge possible in exchange for an admission of guilt or something functionally equivalent to guilt. *See* Andrew Manuel Crespo, *The Hidden Law of Plea Bargaining*, 118 COLUM. L. REV. 1303, 1310–12 (2018).

6. Scott A. Gilbert & Molly Treadway Johnson, *The Federal Judicial Center's 1996 Survey of Guideline Experience*, 9 FED. SENT'G REP. 87, 88–89 (1996); Marc L. Miller, *Domination & Dissatisfaction: Prosecutors as Sentencers*, 56 STAN. L. REV. 1211, 1215, 1219–20 (2004); Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 YALE L.J. 1420, 1422–26 (2008); Besiki Kutateladze et al., *Do Race and Ethnicity Matter in Prosecution? A Review of Empirical Studies*, VERA INST. JUST., 3–4 (June 2012), https://www.vera.org/publications/do-race-and-ethnicity-matter-in-prosecution-a-review-of-empirical-studies.

7. *See* Besiki Kutateladze et al., *Cumulative Disadvantage: Examining Racial and Ethnic Disparity in Prosecution and Sentencing*, 52 CRIMINOLOGY 514, 518, 527-537 (2014).

8. *See* Gail Kellough & Scot Wortley, *Remand for Plea: Bail Decisions and Plea Bargaining as Commensurate Decisions*, 42 BRIT. J. CRIMINOLOGY 186, 194–201 (2002); Besiki Kutateladze et al., *Opening Pandora's Box: How Does Defendant Race Influence Plea Bargaining?*, 33 JUST. Q. 398, 410-419 (2016).

what happened and when is a matter of cobbling together reports from victims, witnesses, police officers, and investigators. In addition, prosecutors must rely on their own past experiences, a reliance that runs numerous risks, including that of small sample size bias. Given these cognitive constraints, prosecutors are liable to rely on stereotypes, such as those that attach to African Americans.[9]

II. PREDICTIVE ANALYTICS IN CRIMINAL JUSTICE

The use of predictive analytics in the law can be bifurcated into two subsets. One involves policing, where what is being predicted is who will commit future crimes.[10] Embedded in this prediction is the question of where those crimes will occur. In theory, these predictions can be used by police departments to allocate resources more efficiently and to make communities safer.

Dozens of police departments around the United States are employing predictive policing.[11] Since 2011, the Los Angeles Police Department ("LAPD") has analyzed data from rap sheets in order to determine how best to utilize police resources.[12] Chicago officials have experimented with an algorithm that predicts which individuals in the city are likely to be involved in a shooting—either as the shooter or as the victim.[13]

The second subset primarily involves recidivism. Here, we have bail decisions in which predictions about who will show up to future court dates are made.[14] Embedded in these predictions is the question of who, if released pretrial, will cause harm (or commit additional crimes).[15] This subset also

9. Decades of research at the nexus of law and psychology have identified stereotypical associations linking blackness with crime, violence, threats, and aggression. *See* Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. PERSONALITY & SOC. PSYCHOL. 1314, 1324-1328 (2002); Jennifer L. Eberhardt et al., *Seeing Black: Race, Crime, and Visual Processing*, 87 J. PERSONALITY & SOC. PSYCHOL. 876, 889-891 (2004); Brian Keith Payne, *Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon*, 81 J. PERSONALITY & SOC. PSYCHOL. 181, 190-191 (2001).

10. *See* Albert Meijer & Martijn Wessels, *Predictive Policing: Review of Benefits and Drawbacks*, INT'L J. PUB. ADMIN. 1, 2-4 (2019).

11. Issie Lapowsky, *How the LAPD uses Data to Predict Crime*, WIRED (May 22, 2018, 5:02 PM), https://www.wired.com/story/los-angeles-police-department-predictive-policing.

12. *Id.*

13. Jeff Asher & Rob Arthur, *Inside the Algorithm That Tries to Predict Gun Violence in Chicago*, N.Y. TIMES: THE UPSHOT (June 13, 2017), https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html.

14. *See*, *e.g., Public Safety Assessment: Risk Factors and Formula*, PUB. SAFETY ASSESSMENT [hereinafter *Risk Factors and Formula*], https://www.psapretrial.org/about/factors (last visited June 6, 2019).

15. *See* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHMENT

includes sentencing, such that judges may receive predictions regarding a defendant's likelihood of recidivating.[16]

The Laura and John Arnold Foundation ("Arnold Foundation") designed its Public Safety Assessment tool ("PSA") to assess the dangerousness of a given defendant.[17] The tool takes into account defendants' age and history of criminal convictions, but it elides race and gender and supposed covariates of race and gender, such as employment background, where a defendant lives, and history of criminal arrests.[18] Risk assessments focusing on recidivism are consulted by sentencing courts.[19] These statistical prediction tools make use of a number of features (factors specific to a defendant) to produce a quantitative output: a score that reflects a defendant's likelihood of engaging in some behavior, such as committing additional crimes or additional violent crimes.[20]

III. AGAINST PREDICTIVE ANALYTICS IN CRIMINAL JUSTICE

Statistical algorithms that have been used for risk assessment have been charged with perpetuating racial bias[21] and have been the subject of litigation.[22] A 2016 report by ProPublica alleging that an algorithm used in Florida was biased against black defendants received nationwide attention.[23] The subsequent debate about whether the algorithm actually was biased against black defendants pivoted on different definitions of fairness, with a specific focus on rates of false positives, true negatives, and related

IN AN ACTUARIAL AGE 1 (2007); Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 61 (2017); Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 808–18 (2014).

16. Melissa Hamilton, *Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law*, 47 ARIZ. ST. L.J. 1, 3 (2015); Anna Maria Barry-Jester et al., *The New Science of Sentencing*, MARSHALL PROJECT (Aug. 4, 2015, 7:15 AM), https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing.

17. *About the PSA*, PUB. SAFETY ASSESSMENT, https://www.psapretrial.org/about (last visited June 6, 2019).

18. *Risk Factors and Formula*, *supra* note 14.

19. Timothy Bakken, *The Continued Failure of Modern Law to Create Fairness and Efficiency: The Presentence Investigation Report and Its Effect on Justice*, 40 N.Y.L. SCH. L. REV. 363, 363–64 (1996); Starr, *supra* note 15, at 803.

20. John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 405–06 (2006).

21. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674, 678 (2016); Jessica M. Eaglin, *Predictive Analytics' Punishment Mismatch*, 14 I/S: J.L. & POL'Y FOR INFO. SOC'Y 87, 102–03 (2017).

22. *See* State v. Loomis, 881 N.W.2d 749, 757–60 (Wis. 2016).

23. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

concepts.[24] Overall, the fear is that, at best, algorithmic decisionmaking perpetuates historical bias; at worst, it exacerbates bias. As one opponent of the LAPD's use of predictive analytics said, "[d]ata is a weapon and will always be used to criminalize black, brown and poor people."[25]

Professor Jessica Eaglin has argued that risk itself is a "malleable and fluid concept"; thus, predictive analytics focused on risk assessment give a spurious stamp of objectivity to a process that is agenda-driven.[26] Furthermore, Professor Eaglin argues that the agenda of these tools is one of increased punishment.

Critics also address the creation of the models. Some argue that the training data is nonrepresentative.[27] Others argue that recidivism is difficult to define[28] and that some jurisdictions are improperly defining it to include arrests, which may be indicative of little beyond police bias.[29] Still others debate which features such models properly should include.[30]

IV. THE IMPORTANCE OF DATA FOR CRIMINAL JUSTICE FAIRNESS

While it is important to question how data is used in criminal justice, the importance of data's role in diminishing racial disparity in incarceration should not be underestimated. First, without robust data collection, we have no way of knowing when similarly-situated defendants are being treated dissimilarly. If we cannot clearly identify racial bias in the different stages of the criminal justice system, then we cannot fix it. And there is still a ways

24. *See* Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks."*, 80 FED. PROB., Sept. 2016, at 38; *see also* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 6 (2014) (calling for predictions that are consistent with normative concepts of fairness).

25. Cindy Chang, *LAPD Officials Defend Predictive Policing as Activists Call for Its End*, L.A. TIMES (July 24, 2018, 8:20 PM), https://www.latimes.com/local/lanow/la-me-lapd-data-policing-20180724-story.html.

26. Eaglin, *supra* note 21, at 105; *see also* Eaglin, *supra* note 15, at 64.

27. *See* Eaglin, *supra* note 15, at 118.

28. Joan Petersilia, *Recidivism*, *in* ENCYCLOPEDIA OF AMERICAN PRISONS 215, 215–16 (Marilyn D. McShane & Frank R. Williams III eds., 1996).

29. *See* Kevin R. Reitz, *Sentencing Facts: Travesties of Real-Offense Sentencing*, 45 STAN. L. REV. 523, 528–35 (1993) (arguing against reliance on unadjudicated conduct at sentencing).

30. *See* Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 BIG DATA 153, 153-162 (2017); Don A. Andrews, *Recidivism Is Predictable and Can Be Influenced: Using Risk Assessments to Reduce Recidivism*, CORRECTIONAL SERV. CAN. (Mar. 5, 2015), https://www.csc-scc.gc.ca/research/forum/e012/12j_e.pdf; Jon Kleinberg et al., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, PROC. OF INNOVATIONS IN THEORETICAL COMPUTER SCI. (forthcoming 2017).

to go before prosecutorial data is properly organized and digitized.[31]

Second, data is essential for collaborative intelligence, which shows significant potential for improving prosecutorial decisionmaking. Prosecutors' offices are in possession of information that can be used to form clear and unbiased baselines: hundreds of thousands of closed casefiles. Using advanced statistical and computer science methods, these casefiles can be used as a corpus from which to build a model that, based on an arresting officer's narrative report and suggested charges, produces a prediction as to how a case would resolve if the defendant were treated race-neutrally. This is a classic machine-learning task: train an algorithm to produce a prediction function that relates case characteristics to case outcomes. This model can then be used to guide prosecutorial decisionmaking to make it more consistent (less variance across attorneys and across time) and less biased.

Algorithms will produce biased outcomes when the training data (the historical record) is biased *and* the algorithm is designed to maximize predictive accuracy. It should be obvious as to why this is the case: if predictive accuracy is the goal and the data is biased, then bias is a feature of the system, not a bug. In other words, bias must be taken into account if the prediction is to be accurate.

This is the reason why, in my research, I do not optimize prediction. My colleagues and I have different goals. Our models are not predictive models but "suggestive" models. One of our primary goals is to remove suspect bias from the model, bringing its suggestions into closer accord with Constitutional mandates for racially equal treatment of criminal defendants by state actors.

Can this be done? It is no easy feat, but researchers around the country are diligently working to build models that correct for suboptimal historical records.[32] Some of these approaches involve a weak version of disparate treatment in which the protected attribute (for example, race) is accessed

31. BESIKI L. KUTATELADZE ET AL., PROSECUTORIAL ATTITUDES, PERSPECTIVES, AND PRIORITIES: INSIGHTS FROM THE INSIDE, MACARTHUR FOUNDATION 2 (2018), https://caj.fiu.edu/news/2018/prosecutorial-attitudes-perspectives-and-priorities-insights-from-the-inside/report-1.pdf; *see also* Andrew Pantazi, *What Makes a Good Prosecutor? A New Study of Melissa Nelson's Office Hopes to Find Out*, FLA. TIMES UNION, https://www.jacksonville.com/news/20180309/what-makes-good-prosecutor-new-study-of-melissa-nelsons-office-hopes-to-find-out (last updated Mar. 12, 2018, 11:18 AM).

32. *See* Alexander Amini et al., *Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure* (2019) (unpublished manuscript), http://www.aies-conference.com/wp-content/papers/main/AIES-19_paper_220.pdf. For another approach at building a non-discriminatory classifier, *see* Irene Chen et al., *Why Is My Classifier Discriminatory?*, *in* 31 ADVANCES IN NEURAL INFO. PROCESSING SYSTEMS 1, 3-9 (2018), http://papers.nips.cc/paper/7613-why-is-my-classifier-discriminatory.pdf.

during model training but omitted during classification.[33] Such approaches build from the recognition, long established in the scholarly community, that not only does blindness not entail fairness,[34] it often is a poor notion of fairness.[35]

Lastly, such models can themselves be used to identify racism that is endemic to the historical record or which emerges in the construction of the model. One strength of machine learning is that it is able to make connections between inputs and outputs that elude human actors. Social science long ago established that the human mind itself is a black box, and human actors have poor insight into their reasons for acting.[36] The black box of human decisionmaking, however, can be unpacked through careful use of statistics. Local-interpretable-model-agnostic explanations,[37] for instance, can be used to identify the aspects of input data on which a trained model relies as it makes its predictions, which should, in turn, offer insight into historical human reliance.[38]

CONCLUSION

When it comes to racial disparities, the U.S. criminal justice system is failing, and it has been failing for many years. In addition, charges of racial bias have been leveled against various organizations that are employing predictive analytics in their legal decisions. Scholars are right to question how data is being used. Past discrimination must not become enshrined in our machines. But movement away from data is also movement away from identification of unequal treatment, and it represents abandonment of the most promising path towards criminal justice fairness. While it is tempting for prosecutors' offices to maintain the status quo and not augment their

33. *See* Zachary C. Lipton et al., *Does Mitigating ML's Impact Disparity Require Treatment Disparity?*, *in* 31 ADVANCES IN NEURAL INFOR. PROCESSING SYSTEMS 1, 9 (2018), https://papers.nips.cc/paper/8035-does-mitigating-mls-impact-disparity-require-treatment-disparity.pdf.

34. Cynthia Dwork et al., *Fairness through Awareness, in* PROCEEDINGS 3RD INNOVATIONS IN THEORETICAL COMPUTER SCI. CONF. 214, 218 (2012), https://dl.acm.org/citation.cfm?id=2090255.

35. Moritz Hardt et al., *Equality of Opportunity in Supervised Learning* 18–19 (Oct. 11, 2016) (unpublished manuscript), https://arxiv.org/pdf/1610.02413.pdf.

36. *See* Richard E. Nisbett & Timothy DeCamp Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes,* 84 PSYCHOL. REV. 231, 251-257 (1977).

37. Introduced by Professors Marco Ribeiro, Sameer Singh, and Carlos Guestrin, "local interpretable model-agnostic explanations," refers to a computer science technique that attempts to explain the predictions of any classifier by learning an interpretable model around the primary prediction. *See* Marco T. Ribeiro et al., *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, ACM 1 (Aug., 2016), https://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf.

38. *See* Michael Chui et al., *What AI Can and Can't Do Yet for Your Business*, MCKINSEY Q., Jan. 2018, at 7, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/what-ai-can-and-cant-do-yet-for-your-business.

processes with data science, this would be a mistake. Collaborative intelligence has the potential to render prosecutorial decisionmaking more consistent, fair, and efficient.