SCIENCE AND ETHICS OF ALGORITHMS IN THE COURTROOM

Kia Rahnama†

Abstract

This Article analyzes the societal and cultural impacts of greater reliance on the use of algorithms in the courtroom. Big-data analytics and algorithms are beginning to play a large role in influencing judges' sentencing and criminal enforcement decisions. This Article addresses this shift toward greater acceptance of algorithms as models for risk-assessment and criminal forecasting within the context of moral and social movements that have shaped the American justice system's current approach to punishment and rehabilitation. By reviewing salient problems of scientific uncertainty that accompany the use of these models and algorithms, the Article calls into question the proposition that greater reliance on algorithms in the courtroom can lead to a more objective and fair criminal sentencing regime. Far from liberating the society from the biases and prejudices that might pollute judges' decision-making process, these tools can intensify, while simultaneously concealing, entrenched cultural biases that preexist in the society. Using common themes from the field of Science and Technology Studies (STS), including boundary-work analysis and Public Understanding of Science (PUS), this Article highlights unique technical characteristics of big-data analytics and algorithms that feed into undesirable and deeply-held values and beliefs. This Article draws attention to specific gaps in technical understanding of algorithmic thinking, such as the black box of algorithms, that can have discordant impact on communicating uncertainty to the populace and reduce accountability and transparency in regulating the use of algorithms. This Article also provides specific policy proposals that can ameliorate the adverse social and cultural effects of incorporating algorithms into the courtroom. The discussion of policy proposals borrows from the STS literature on public participation in science and encourages adoption of a policy that incorporates diverse voices from political actors, most affected communities, and the offenders themselves. This Article was accepted into the ST Global Conference, and I presented the Article on March 23, 2018.

† J.D./MA 2018, George Washington University Law School and the Elliott School of International Affairs. I would like to thank the dedicated staff and editors of the Journal of Law, Technology & Policy. I am grateful to Professor Allison MacFarlane for her enlightening lectures on issues of science and technology policy and her feedback and guidance with this Article. Finally, I would like to thank Natalie Tarasar, Masha Rahnama, Arash Rahnama, Ardeshir Rahnama, and Mozhdeh Shahin for their endless support.

TABLE OF CONTENTS

I. INTRODUCTION

Big-data analytics is growing root in many aspects of the daily life in the developed world.[1] It is being used to collect and integrate more accurate patient-centric information for medical centers and researchers,[2] to systematically review swaths of historical and archival financial information to help institutions predict financial behavior,[3] and to organize and sort information on many of the most widely used social network platforms.[4] Big-data analytics assists decision makers by making special use of the brute computational force of today's computers in detecting underlying patterns in large repositories of data, patterns that might be inaccessible to humans through the application of traditional analytical methods but can be useful in forecasting future events.[5] Not surprisingly, public officials have shown openness in utilizing this revolution in computer science and technology with the intention to ameliorate an unparalleled malady facing the American criminal justice system. The United States visibly suffers from abnormally high levels of incarceration,[6] sky-rocketing prison costs,[7] and high rates of recidivism and enforcement

1. Naveen Joshi, *Big Data Is Reinventing Our Society*, ALLERIN (Oct. 12, 2016), https://www.allerin.com/blog/big-data-is-reinventing-our-society.

2. Bob Violino, *Optimizing Organ Donation: When Big Data and Analytics Helps Save Lives*, ZDNET (Dec. 6, 2017, 1:05 PM), http://www.zdnet.com/article/optimizing-organ-donation-when-big-data-and-analytics-helps-save-lives.

3. John Edwards, *What Is Predictive Analytics? Transforming Data Into Future Insights*, CIO (May 16, 2018, 3:00 AM), https://www.cio.com/article/3273114/predictive-analytics/what-is-predictive-analytics-transforming-data-into-future-insights.html.

4. Alexis C. Madrigal, *Behind the Machine's Back: How Social Media Users Avoid Getting Turned Into Big Data*, THE ATLANTIC (Apr. 14, 2014), https://www.theatlantic.com/technology/archive/2014/04/behind-the-machines-back-how-social-media-users-avoid-getting-turned-into-big-data/360416.

5. Keith Collins, *Big Data Demands Big Computing*, INFO. WK. (Mar. 2, 2015, 9:36 AM), https://www.informationweek.com/big-data/big-data-analytics/big-data-demands-big-computing/a/d-id/1319273.

6. Michelle Ye Hee Lee, *Yes, U.S. Locks People Up at a Higher Rate Than Any Other Country*, WASH. POST (July 7, 2015), https://www.washingtonpost.com/news/fact-checker/wp/2015/07/07/yes-u-s-locks-people-up-at-a-higher-rate-than-any-other-country/?utm_term=.6d79f8ea1506.

7. Matt Ferner, *The Full Cost of Incarceration in the U.S. is Over $1 Trillion, Study Finds*, HUFFINGTON POST (Sept. 13, 2016, 5:01 PM), https://www.huffingtonpost.com/entry/mass-incarceration-cost_us_57d82d99e4b09d7a687fde21.

inequality.[8] The consensus on the need for reform is palpable, and both sides of the political debate in America can seek refuge under the banner of an objective and science-led approach to reform, which promises meaningful change, yet circumvents many of the ethical and moral debates about criminal enforcement that divide the two sides.[9]

The big data revolution can be incorporated into the science of sentencing and criminal enforcement through the use of algorithms and risk-assessment tools.[10] These risk-assessment tools can match the information obtained from individual criminal defendants with the patterns observed among past offenders with similar background and make probabilistic judgments about defendants' future conduct.[11] The problem arises when these patterns themselves arise from foundationally undesirable conditions that breed discrimination and disadvantage, rendering the pool of data accessible to the algorithms polluted with bias.[12] This and many other complex aspects of algorithmic-based decision-making invite caution against systematic incorporation and adoption of machines into the courtroom.[13] Critical analysis of the use of algorithms in the criminal justice system will inevitably trigger a discussion about reoccurring themes in science and technology studies, themes that cast questions on the presumption that brute computational force in conjunction with a scientific approach to quantification can yield appropriately fair and objective results.[14] This Article applies some of the common criticisms in the field of science and technology studies to the use of algorithmic risk-assessment tools in the criminal system and proposes solutions that can mediate some of the societal concerns over adoption of this new technology.

II. BACKGROUND ON AMERICAN CRIMINAL SENTENCING PROCESS

In the nineteenth and early twentieth century the American legal system largely endorsed a "rehabilitative medical model" of criminal sentencing.[15] Under this system, state and federal legislatures would often only adopt maximum sentences for a range of criminal offenses but leave state and federal judges with "unfettered discretion" to impose any sentences within the context of the broad statutory range provided.[16] This system emphasized erasing the

8. *See* MICHELLE ALEXANDER, THE NEW JIM CROW: INCARCERATION IN THE AGE OF COLORBLINDNESS (The New Press 2010) (discussing incarceration inequalities).

9. Shalia Dewan & Carle Hulse, *Republicans and Democrats Cannot Agree on Absolutely Anything. Except This.*, N.Y. TIMES (Nov. 14, 2018), https://www.nytimes.com/2018/11/14/us/prison-reform-bill-republicans-democrats.html.

10. Laurel Eckhouse, *Big Data May Be Reinforcing Racial Bias in the Criminal Justice System*, WASH. POST (Feb. 10, 2017), https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d_story.html.

11. *Id.*

12. *Id.*

13. *Id.*

14. *Id.*

15. Lynne Woodruff, *The Evolution of Prison Design and the Rise of the Direct Supervision Model*, LEXIPOL (Mar. 17, 2017), https://www.lexipol.com/resources/blog/the-evolution-of-prison-design-and-the-rise-of-the-direct-supervision-model.

16. Douglas A. Berman, *Forward: Beyond Blakely and Booker: Pondering Modern Sentencing Process*, 95 J. CRIM. L. & CRIMINOLOGY 654, 654 (2005).

psychological and occupational characteristics that had hindered the criminals' integration into the society in the first place.[17] Judges also shared with parole officials a broad power in assigning flexible prison release dates that corresponded with the rehabilitative progress of each suspect.[18] The presumption underlying this system was that judges and parole officials had "unique insight and expertise" in determining the nature of consequences that would flow from each type of punishment imposed on the defendants.[19]

Two significant factors led to a foundational shift in the American criminal procedure in the decade between 1960 to 1970.[20] First, an increasing crime rate, and the modern media's new role in covering crime-ridden communities, which intensified the social impact of this trend, shifted the policy makers' focus to a more certainty-based deterrence approach.[21] Second, empirical review of the rehabilitative system revealed undue inconsistencies in lengths and types of sentences assigned to similar crimes in different parts of the country—often accompanied with evidence linking the inconsistency to race, gender, or socioeconomic discrimination by the judges.[22] In order to arrive at a more consistent, firm, and predictable system of sentencing that could better deter future crimes, American law broadly endorsed a movement toward reliance on pre-determined and uniform sentencing structures drafted and imposed by the legislature.[23] This task invariably dictated a shift in prioritizing different types of knowledge.[24] Whereas the judge-centered sentencing regime valued the judges' impressionistic and subjective assessment of the information available about individual offenders (information that would have been produced in cooperation with the prosecutors who were charged with conducting pre-sentence investigations on criminals and assembling information relevant to sentencing), the legislation-centered sentencing focused on the legislators' knowledge of the social perception of the gravity of different crimes.[25] This shift toward legislatively-mandated, uniform guidelines reflected increasing societal concern at this time over the second-degree effects of criminal activity on the society, such as high financial cost of managing the prison system and issues of public safety related to recidivism rates.[26] As a result of this intellectual shift, many states established sentencing commissions charged with designing presumptive sentencing ranges, and in the 1980s, the federal government created

17. *See id.* at 655 (noting some of the factors that criminal justice researchers started focusing on, regarding sentencing policy).

18. *Id.*

19. *Id.*

20. *Id.*

21. Berman, *supra* note 16, at 655.

22. *Id.*

23. *Id.* at 657.

24. *See id.* at 654–59 (noting the differing priorities of the discretionary sentencing model and those who advocated for greater uniformity in sentencing).

25. *See, e.g., id.* at 664 (explaining the legislative intent behind Pennsylvania's Mandatory Minimum Sentencing Act).

26. *See generally* Berman, *supra* note 16, at 655 (noting criticisms of the discretionary model of sentencing, because of the social costs associated with it).

the U.S. Sentencing Commission and developed uniform guidelines for federal sentencing.[27]

Today, decades after this transition, criminal sentencing in America is more controlled, predictable, and structured than any time in its past history.[28] Although judges are encouraged by most sentencing commissions to pay special attention to fact-finding and opportunities for parties to argue why sentencing outcomes should diverge from the pre-established guidelines issued by the legislature, the sentencing guidelines have perceptible effects on ultimate decisions.[29] Against the background of this dialogue between the legislature and the judiciary over arriving at more predictable and consistent sentencing patterns, the rise of risk-assessment algorithms and models opened a new avenue for even further progression toward systematically uniform approaches to criminal sentencing.[30]

III. MODELS AND ALGORITHMS IN CRIMINAL JUSTICE

The use of models and algorithms in criminal sentencing is a product of the recent shift in focus toward evidence-based practices.[31] This approach contends that some isolated, yet uniform, factors contribute to the risk of crime and recidivism by influencing the behavior patterns of a large portion of the population in a similar and predictable way.[32] Models and algorithms can process and review vast amounts of data in order to identify and trace these factors and signify their weight and force in causing criminal behavior.[33] Algorithms refer to descriptions of the methods by which a task is to be accomplished; they are program behaviors by which the programmer or the developer of the system intends the machine to follow.[34] This program behavior is often communicated to the machine through the use of coding that expresses "the promise of an algorithm."[35] Models and algorithms, when used to conduct risk-assessment, can provide risk probabilities by mechanically combining the inputs in demographic data-sets into classifications and categories and using the relationship between these categories to arrive at probability figures about relative frequency of certain events.[36]

27. *Id.* at 657–60.

28. *See generally* Alyssa M. Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L.R. 303, 305 (2017) [hereinafter Carlson] (noting the moves that the U.S. has made towards developing a more objective and predictable sentencing regime).

29. *See generally* Berman, *supra* note 16, at 659 (discussing the manner in which substantive sentencing laws exert an influence over judges when it comes to sentencing, but often leave them with discretion regarding the process that judges assess the facts of the case).

30. *See* Carlson, *supra* note 28, at 327 (noting the efforts in Ohio to use a statewide risk assessment system to improve sentencing consistency and to facilitate communication across criminal justice agencies).

31. *See id.* at 305 (describing the increasing reliance on risk assessment systems in sentencing decisions, and how this new approach differs from the "clinical model" of sentencing decisions).

32. *Id.*

33. *See id.* at 309 (explaining that the most widespread predictive tools are "questionnaires that assign points based on factors such as demographics, family backgrounds, and criminal history.").

34. Paul Dourish, *Algorithms and Their Others: Algorithmic Culture in Context*, BIG DATA & SOC'Y 1, 3 (2016).

35. *Id.* at 3–4.

36. *Id.*

In the most sophisticated risk-assessment algorithms, machine learning techniques are used to strengthen the computational power of machines even further.[37] Machine learning technology provides the machines with the opportunity to tweak the operational parameters provided by the original programmer and fit the decision-making rules to the manner best suited for reaching the algorithm's goals.[38] Instead of relying on detailed code, machine learning algorithms allow the machines to train themselves on how to best process the data; this will sometimes enable the machine to modify the original rules provided to it on how to classify or treat the data.[39] Machines operating under these algorithms dynamically adjust to new data by relying on patterns they observe in big data.[40] Machine learning algorithms are allowed to deviate from the programmers' definitive data processing procedures if they learn better ways to arrive at the outcome than the programmer intended.[41] Big-data algorithms that use machine learning, therefore, produce knowns through unknown means.[42] Machine learning algorithms' ability to modify their own processing structure means that, in arriving at conclusions, they might rely on assumptions about relationships between different categories of data that may remain hidden even to the systems' designers.[43] When the algorithm's rationale for a certain finding is inaccessible to even the programmer of the algorithm, the condition is referred to as the "black box" of algorithm.[44]

Today, risk-assessment models are used in a wide range of decisions affecting defendants, from sentencing and parole determinations, to sketching the conditions of supervised release, reentry services, and judgments regarding probation.[45] Given the sophisticated structure of the modern algorithms that are used to create risk-assessment tools, private companies' services are often procured by state and federal bodies to further hone the efficacy of the models.[46] Early risk-assessment models claimed to accurately predict success or failure of parole based on relatively little information, with the models' input ranging from 7 to 21 factors about the defendants.[47] Today, over 60 different types of risk-assessment tools are currently in use in courthouses across the country.[48] With some variety, these models assign points to defendants' profiles based on factors that are interlinked with demographics, family background, and criminal history, putting specific emphasis on "antisocial attitudes, antisocial associates,

37. Brent Daniel Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, BIG DATA & SOC'Y 1, 1 (2016).

38. *Id.* at 2.

39. *Id.* at 6.

40. DANIELLE KEHL ET AL., HARV. L. SCH., ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM: ASSESSING THE USE OF RISK ASSESSMENTS IN SENTENCING 6 (2017), https://dash.harvard.edu/bitstream/handle/1/33746041/2017-07_responsivecommunities_2.pdf?sequence=1.

41. Dourish, *supra* note 34, at 7.

42. *Id.*

43. Mittelstadt et al., *supra* note 37, at 6.

44. *Id.*

45. Carlson, *supra* note 28, at 308.

46. *Id.* at 306.

47. *Id.* at 307–09.

48. *Id.* at 309.

antisocial personalities, and criminal history."[49] Substance abuse, family characteristics, education, employment, and lack of prosocial leisure or recreation are also often considered to be of importance in a large number of models.[50] The literature currently supports the proposition that the predictive-model approach prioritizes information about sex, age, and prior criminal history at the highest level in assigning individual predictive scores.[51]

The Justice Department's National Institute of Corrections currently promotes the use of predictive algorithms for all phases of criminal enforcement, and "an overwhelming majority of correction agencies nationwide routinely utilize" the models.[52] So far, the empirical evidence comparing the track-record of well-designed risk assessment tools in predicting behavior to predictions from unaided expert opinion shows a positive trend in the contributions these tools can make to human decision-making. In general, the risk-assessment tools tend to outperform expert opinion by about 10 percent.[53] As market and government incentives further encourage exploration and utilization of new and sophisticated model-based approaches to criminal sentencing,[54] it is worth considering whether an overconfidence in the predictive power of big-data analytics and algorithms may leave society vulnerable to systematic epistemological errors.

A. Algorithms and Scientific Objectivity

Algorithms can be perceived as appropriate tools for guiding the state's retributivist actions over its citizens because decisions based on algorithmic findings can appear more detached and objective.[55] This popular perception is in conflict with one of the central presumptions in the field of science and technology studies which calls into question the assumption that tools of scientific discourse created by a value-laden human culture could ever be purely objective, dispassionate, and disassociated from the values endorsed by social process.[56] Under this paradigm, it is important to remember that the operational parameters of algorithms are designed by developers and programmers with specific desires, outcomes, and biases in mind and that these desired outcomes invariably privilege some values and interests over the others.[57] As much as the discourse favoring the use of algorithms might highlight the necessity of avoiding the worst cases of human bias reflected in judges' decision-making process, the underlying truth that algorithms will also be designed and created by people who inevitably hold value-laden presumptions and intuitions is

---

49. *Id.*

50. *Id.*

51. *Id.* at 310.

52. *Id.* at 313.

53. 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.

54. KEHL ET AL., *supra* note 40, at 7–8.

55. *See* Rumman Chowdhury & Narendra Mulani, *Auditing Algorithms for Bias*, HARV. BUS. REV. (Oct. 24, 2018), https://hbr.org/2018/10/auditing-algorithms-for-bias ("However objective we may intend our technology to be, it is ultimately influenced by the people who build it and the data that feeds it.").

56. *See generally* THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 128 (Chi. Univ. Press, 2d ed. 1962) (discussing assumptions surrounding the objectivity of technological discourse).

57. Mittelstadt et al., *supra* note 37, at 7.

inescapable.[58] Much like the human involvement in other areas of scientific thought, in coding and algorithm design too, "the values of the author wittingly or not, are frozen into the code, effectively institutionalizing those values."[59]

An example of a discretionary decision by developers of algorithms, and one that will inevitably involve value trade-offs, is the critical decision on how to fine-tune the algorithm with regards to its receptiveness to false positives and false negatives; should the designer aim to adjust the model to be particularly sensitive to scenarios where potential reoffenders are erroneously labeled as low-risk offenders, or should the algorithm instead be sensitive to the scenarios where those who might not be reoffenders are erroneously labeled as high-risk reoffenders.[60] The unavoidable task of prioritizing the algorithm's approach to uncertainty in such scenarios is testament to an inherent inadequacy of the models.[61] The likely outcome in such scenarios might itself be influenced by the political incentives of different governmental bodies who fund the development of algorithms.[62] Given the right alignment of political incentives, and presumably when public input has played a role in the design of the algorithm, models can reflect the community's priorities and have salutary—yet still subjective—effects.[63] In Charlotte, North Carolina, with the intention of curtailing human bias that could have been contributing to the city's burdensome bail system, transitioned to relying on a risk-assessment tool that was specifically fine-tuned to remedy that human bias.[64] The model allowed Charlotte to cut its jail population by 20 percent without any perceptible increase in the crime rate.[65] It is therefore possible that algorithms that shift the ultimate decision-making power from the judges to algorithms might do a better job at ensuring that the concern over the high fiscal cost of incarceration, which is often important to the legislature and not the independent judiciary, is better reflected in the states' criminal enforcement decisions.[66]

Algorithms' use of quantification and categorization will also lead to biases against considering intangible characteristics such as community ties and in favor of what could be more easily put in numbers, such as employment status, age, or history of drug abuse.[67] Numbers and simple categories have the benefit of serving as shortcuts that reduce the risk of misinterpreting new data or

58. Matthias Spielkamp, *Inspecting Algorithms for Bias*, MIT TECH. REV. (June 12, 2017), https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias.

59. Mittelstadt et al., *supra* note 37, at 7.

60. *See id.* at 10 (discussing the decision-making structure with respect to confidence intervals and the resulting impact it has on false positives, false negatives, and spurious correlations); Spielkamp, *supra* note 58 (discussing automated decision-making systems).

61. *See* Mittelstadt et al., *supra* note 37, at 10 (discussing traceability leading to moral responsibility).

62. Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y 7 (2016).

63. *Id.*

64. Joe Killian, *State Panel Commences Examination of NC's Cash Bail System*, NC POL'Y WATCH (Oct. 18, 2018), http://www.ncpolicywatch.com/2018/10/18/state-panel-commences-examination-of-ncs-cash-bail-system.

65. Shaila Dewan, *Judges Replacing Conjecture with Formula for Bail*, N.Y. TIMES (June 26, 2015), https://www.nytimes.com/2015/06/27/us/turning-the-granting-of-bail-into-a-science.html.

66. Barry-Jester et al., *supra* note 53.

67. Mike Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, 41 SCI. TECH. & HUM. VALUES 1, 11 (2016).

situations, but an emphasis on these shortcuts also tends to marginalize relevant information that requires deeper exploration, and yet might not be easily conducive to quantification or categorization—the format most useful for algorithms.[68] This problem closely resembles measurement biases in the context of cost/benefit analysis.[69] One common criticism of utilitarian cost/benefit analysis in the context of environmental regulation highlights the bias that arises from exclusion of subjective categories such as "the value of natural objects and of historical and cultural monuments and practices" from utilitarian calculations because they cannot easily be represented in dollar terms.[70] To counter the similar measurement bias in the case of algorithms, some states attempt to complement the data-set with a face-to-face interview with offenders.[71] These interviews move beyond considering a narrowly constructed list of numbers and categories and try to provide algorithms with a more wholistic data-set.[72]

In addition to the practical problems associated with designing and programming algorithms that produce truly scientifically objective decisions, the societal decision to rely on predictive-analysis tools and algorithms itself reveals a subjective stance.[73] A public policy approach to criminal justice reform that emphasizes funding of scientific tools that facilitate punishment and deterrence implicitly endorses certain theories about criminal behavior at the expense of others. As an example, adherence to the theory of "selective incapacitation," which ascribes the high rate of criminal activity in the country to a "small subset of repeat offenders," and consequently recommends incapacitating that small group as a way to handle high crime rates, would specially favor the use of tools that directly or indirectly bring law enforcement closer to identifying that small subgroup.[74]

New technologies also impact societies by introducing tools that strengthen preexisting societal biases and perpetuate arbitrary and subjective decisions on what type of knowledge to pursue and what type of knowledge to ignore: "the often invisible problem of how we decide what it is we are going to try to know, and what, as a consequence, we decide, even if by benign neglect, we are not going to know."[75] Consider one potential impact of pervasive use of predictive algorithms in a society that perceives criminal behavior as an extension of each individual's immutable characteristics as opposed to a direct byproduct of externally imposed societal conditions. In calculating the risk of recidivism, most models have to match information about currently-in-place supervision and rehabilitative treatments provided by the state with the likelihood that these services will reduce certain offender's risk of recidivism.[76] To do this, the model

---

68. *Id.*
69. *See generally* Susan Rose-Ackerman, *The Limits of Cost/Benefit Analysis When Disasters Loom*, 7 GLOBAL POL'Y 56, 58 (2016) (discussing measurement biases).
70. *Id.*
71. Dewan, *supra* note 65.
72. *Id.*
73. Carlson, *supra* note 28, at 304.
74. *Id.*
75. Daniel Sarewitz, *Normal Science and Limits on Knowledge: What We Seek to Know, What We Choose Not to Know, What We Don't Bother Knowing*, 77 SOC. RES. 997, 997 (2010).
76. Carlson, *supra* note 28, at 307–08.

will attempt to assess how successful the rehabilitative programs and services provided by the state, and currently available to the offender, will be when they are used to treat the specific defendant's cognitive and behavioral characteristics.[77] Therefore, the probabilistic predictions made by the algorithm in each scenario will only predict how the defendant will fare given the states' *current* approach to rehabilitation and treatment, completely avoiding the larger discussion over how these programs *could be* improved to lower recidivism rates for defendants with isolated cognitive and behavioral characteristics.

It is easy to see that under an alternative framing, the predictions made by the algorithm are revealing more about the conditions of the states' rehabilitative apparatus as opposed to the criminals' immutable characteristics. Incorporation of technological tools that shift the society's empirical focus to predicting seemingly untreatable risky behavior by offenders, as opposed to focusing on improving the state's role in rehabilitating crime, legitimizes one understanding of criminal behavior at the expense of the other. The type of knowledge we choose to create can in-turn shape the democratic debates in the society.[78] Some impacts of this bias can already be observed in public policy debates in America. For example, a robust shift toward reliance on uniform sentences, and development of scientific tools to achieve uniform sentences, has drastically reduced discussion over other fundamental issues in sentencing such as burdens of proofs, notice to parties and evidentiary rules and hearing processes.[79] Similarly, a focus on actuarial instruments that assess the risk of reoffending in first-offenders by analyzing the traits first-offenders might share with repeat offenders already in prison, conveniently shifts the attention away from studying prison conditions that might have transformed those offenders into repeat offenders.[80]

Many risk-assessment instruments used by various enforcement agencies across the country also routinely include in their data-gathering evaluations questions that take a presumptive and subjective attitude toward serious ethnic and social dilemmas.[81] One algorithm asks defendants to share whether they agree or disagree with the following statement: "A hungry person has a right to steal."[82] The perennial ethical and moral quandaries embedded in such a statement are undoubtedly far harder to resolve than the simple information about the correlative power of different responses to that statement.[83] The latter can easily be assessed through the use of normal science, whereas the former invites lengthy deliberation from experts in a wide range of scientific areas.[84] The decision to pursue knowledge about the specific degree of correlation resulting from different responses to that statement effectively acts to limit knowledge about whether the responsibility for the harmful correlation between

77. *Id.*
78. Sarewitz, *supra* note 75, at 998.
79. Berman, *supra* note 16, at 654.
80. Carlson, *supra* note 28, at 313–16.
81. *See id.* at 311 (noting the nature of questions used in algorithms).
82. *Id.*
83. *Id.*
84. *Id.*

hunger and crime lies with the state or the potential criminal.[85] Over the long run, it can be anticipated that utilization of models that take such answers for granted will shift the democratic debate toward assessing the accuracy of observed correlation between social conditions such as hunger and criminal behavior, and significantly limit democratic debate over deeper deliberation about the state's responsibility in creating those social conditions in the first place.[86]

A stronger adherence to the use of algorithms that focus on punishment will invariably define future "knowledge-creation paths" that divert from exploration of ethical and moral questions about the relationship between the state and offenders—questions that are not as easily solved by algorithms.[87] Attachment to pursuing this kind of scientific and technological knowledge further strengthens institutional arrangements that already operate on one set of subjective understandings about the relationship between the state and the offenders.[88] The state's approach to retributivist justice and what kind of scientific and technological knowledge to pursue (and consequently, what kind of scientific and technological tools to use) defines what social circumstances should matter the most and diminishes further inquiry into many complex relationships between criminal behavior and historical, political, and cultural factors.[89] As opposed to reflecting scientific objectivity, the algorithm becomes a mere embodiment of already "deeply held cultural values and beliefs."[90] Scientists and academics will soon find themselves occupied with satisfying the regulators by providing reliable methods of assessment and measurement without questioning any of the latent social or political assumptions about agency, causality, responsibility, and deterrence in the criminal context.[91]

Models and algorithms, in wearing the cloak of objectivity, can simply perpetuate underlying unequal and unjust conditions that preexist in the society.[92] One of the most commonly used risk-assessment tools in the country, Correctional Offender Management Profiling for Alternative Sanctions

85. *Id.*

86. *See generally id.* (describing various algorithm models).

87. Sarewitz, *supra* note 75, at 1001.

88. *Id.*

89. *See Retributive Justice*, STAN. ENCYCLOPEDIA PHIL. (June 18, 2014), https://plato.stanford.edu/entries/justice-retributive (describing the concept of retributive justice).

90. Sheila Jasanoff, *The Songlines of Risk*, 8 ENVTL. VALUES 135, 135 (1999).

91. *Id.* at 150.

92. Eric Holder, Attorney General, U.S. Dep't of Justice, Nat. Ass'n of Criminal Def. Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference, (Aug. 1, 2014), https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th (describing that in his speech at the 57th Annual Meeting of the National Association of Criminal Defense Lawyers, former U.S. Attorney General Eric Holder, reiterated this concern: "[a]lthough [risk assessment tools] were crafted with the best of intentions, I am concerned that they may inadvertently undermine our efforts to ensure individualized and equal justice. By basing sentencing decisions on static factors and immutable characteristics—like the defendant's education level, socioeconomic background, or neighborhood—they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.").

(COMPAS), has already been shown to have disparate impacts on black defendants.[93]

B. Boundary-Ordering Devices and Sentencing Authority

Given the enormous human cost associated with criminal sentencing decisions and the existence of a shared-responsibility regime between a number of different social entities that play a role in criminal enforcement (ranging from political actors in three separate branches of government to academics, civil society advocates, and even the private market), it is not surprising that resort to boundary-work is likely to have a perceptible presence in the discourse over the use of new technologies in courts.[94] Boundary-work can determine how different players in different institutions with influence over criminal sentencing will use rhetorical boundaries to frame the impact of the specific type of knowledge instantiated in their institutions.[95] Uncertainties and flaws in applying algorithms to the criminal justice system can lead to wrongful convictions, and the experts' discussion of uncertainty will, therefore, have direct bearing on who might be held accountable and responsible for harms caused by faulty algorithms.[96]

Boundary-work has already had a preeminent role in how different institutional participants in the criminal justice system have framed the relationship between the specific type of knowledge they possess and public policy actions.[97] As an example, as criticisms amassed during the first phase of criminal justice reform in America that the discretionary powers granted to judges in sentencing criminals had resulted in great uncertainty, many advocates who wanted to preserve the judicial branch's active role in sentencing policy engaged in "displacement of uncertainty" to shift the responsibility for a major unresolved uncertainty in sentencing to other agents and policy domains.[98] In his 1973 book, *Criminal Sentences: Law Without Order*, Federal District Judge Marvin E. Frankel shifted the responsibility for the unpredictability embedded in the criminal sentencing system to the legislators, contending that "sentencing process is necessarily arbitrary because [of the] irrational penalty provisions prescribed by legislators."[99] This type of boundary-work allowed the judicial branch to argue that insufficiencies in criminal sentencing policies did not result from uncertainty or unreliability in judicial knowledge; rather it arose because of insufficiency in the knowledge possessed by a separate branch of

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

94. Thomas F. Gieryn, *Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists*, 48 AM. SOC. REV. 781, 781–82 (1983).

95. *See generally id.* (discussing how boundary-work impacts several different institutions).

96. *See generally* Carlson, *supra* note 28, at 307–08 (discussing the use of algorithms in criminal sentencing).

97. Simon Shackley & Brian Wynne, *Representing Uncertainty in Global Climate Change Science and Policy: Boundary-Ordering Devices and Authority*, 21 SCI. TECH. & HUM. VALUES 275, 290–92.

98. *Id.*

99. MARVIN E. FRANKEL, CRIMINAL SENTENCES: LAW WITHOUT ORDER (1973).

government.[100] This incentive for engaging in boundary-work is likely to manifest itself in the discourse over the scientific uncertainty of the use of algorithms too.

When algorithms are used in criminal sentencing three different institutions play a part in injecting their own expert knowledge into the public policy process. Legislative knowledge determines whether algorithms should be used in the court system.[101] The developers of algorithms in the private market then use their expert knowledge to design and develop the algorithms, and finally, judges utilize judicial expert knowledge to determine the importance of the role algorithms will play in the courtroom.[102] Although judges currently reserve broad discretion in using the models, discourse evidence suggests that some judges feel comfortable being on the record as admitting that they have supplanted their own subjective expert assessments for the expert knowledge that can be gleaned from consulting the risk-assessment algorithms.[103] This continued adherence to "displacement of uncertainty" by the judges moves responsibility for potential uncertainty in the models and the sentencing decision from the judges to the experts who designed the algorithms.[104] This type of boundary-work can, to some extent, secure and protect the authority of judge-made knowledge by distinguishing it from the knowledge imbedded in the algorithms by the scientific experts and protect the judges from public scrutiny that might arise later on from detection of flaws in algorithms used in the courtroom.[105]

It could be conceivable that the developers of algorithms will also participate in boundary-work to protect and secure the authority of their own expert knowledge. For example, if algorithms are found to perpetuate discriminatory conditions, algorithm developers can transform uncertainty about designing color-blind algorithms by linking discriminatory results of the algorithm to the misguided information embedded in the underlying data fed into the algorithm. This boundary-work by algorithm developers will rest on the presumption that algorithms are only as reliable and color-blind as the underlying data they are based on. If discriminatory results arise out of algorithmic decisions, one safe avenue for auditing uncertainty would be for the expert programmers to openly admit that the repository of big data used by the algorithms contain questionable presumptions that arise from preexisting discriminatory practices by the law enforcement authorities in the society.[106] This shifts the blame to investigators and officers whose discriminatory criminal enforcement practices produce a corrupt data-set of demographic information;

100. *See* James R. Thompson & Gary L. Starkman, *Criminal Sentences: Law Without Order by Marvin E. Frankel*, 74 COLUM. L. REV. 152, 152–58 (1974) (discussing Judge Frankel's views on insufficiencies in criminal sentencing policies).

101. Carlson, *supra* note 28, at 307–08.

102. *Id.* at 309–11.

103. *Id.* at 311, 313, 319–21.

104. *See id.* (discussing the use of such algorithms).

105. *See id.* (discussing the use of algorithms).

106. *See* Danielle K. Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. U. L. REV. 1, 4 (2014) (discussing how humans program predictive algorithms, and therefore, their biases can be embedded into the software).

algorithms only produce discriminatory results because they operate based on the data-set that itself reflects information produced as result of undesired discriminatory practices in the society.[107] Although the public might find itself rightly concerned with the validity and neutrality of the underlying data consumed by the algorithms and the potential that algorithms might perpetuate discriminatory results, the algorithm developers use of transformation of uncertainty will deflect responsibility for the uncertainties associated with the presuppositions embedded in the data used by the algorithms, instead framing their boundary of expert knowledge to be limited to the uncertainties associated with measurement errors and other technical aspects of algorithmic process, "puzzles that [are] amenable to solution by [] routine techniques and methods."[108]

Furthermore, as mentioned earlier, the latest advents in machine-learning technology create the potential for black-box algorithms, where machines have the capacity to gain "informational advantage" over their programmers and operators and arrive at conclusions that cannot be easily understood or articulated by the designers through merely analyzing the inputs or reviewing the code.[109] Such scenarios can conceivably incentivize the legislators who endorse the use of algorithms in the courtroom, or the experts who design and sell the algorithms, to resort to "scheduling into the future" when they are criticized for harms caused by uncertainties in the algorithm.[110] These parties can deflect responsibility over harms caused by uncertainty in the science of algorithms by identifying "when and how the key uncertainties will be reduced," projecting the disturbing dimensions of the ungraspable science of algorithms into the future "when better data, computing power, [or] theories," will emerge to help the developers solve the black box problem.[111]

C. Communicating Uncertainty and Public Understanding of Science (PUS)

Communicating uncertainty is a critical element of expert-driven policymaking. Decisionmakers have the right, and often the desire, to inquire about uncertainties embedded in expert-knowledge.[112] Yet, at least two distinct social conditions associated with the use of algorithms in criminal sentencing create dilemmas that impact communication of the uncertainty existing in experts' knowledge.[113] The first of these conditions involves the significant trade-off that arises from public-private partnerships in designing and using up-to-date algorithms.[114]

---

107. *Id.*
108. Shackley & Wynne, *supra* note 97, at 285.
109. Mittelstadt et al., *supra* note 37, at 6–7.
110. Shackley & Wynne, *supra* note 97, at 287–90.
111. *Id.* at 287.
112. Baruch Fischhoff, *Communicating Uncertainty Fulfilling the Duty to Inform*, 28 ISSUES IN SCI. & TECH. 63, 64 (2012).
113. *See generally* Carlson, *supra* note 28, at 319–21 (describing the application of criminal sentencing algorithms and the lack of transparency).
114. *Id.*

Distinct from many of the other controversial uses of science and technology in regulating the society, the specific context of science and technology in the courtroom naturally exposes the government to higher risk of liability.[115] Because the application of science and technology in this context has a first-degree impact on the lives and fortunes of many citizens, potential for legal challenges that call the government's approach into question are relatively higher.[116] To take the application of government-funded science and technology to climate change regulation as an example, the courts have largely refused to accept a private right of action for citizens in challenging broad governmental scientific policies that affect the society as a whole, finding the political process to be a more suitable remedy for correcting political uses of science and technology in that context.[117] Conversely, criminal defendants who have been directly subjected to the abuses of scientific and predictive models endorsed by the government in criminal sentencing have routinely used the courts to hold the state responsible for its decision to rely on certain models.[118]

The litigious side effects of using algorithms in criminal sentencing raises unique problems for transparency and dialogue.[119] Private providers of models and algorithms are likely to contend that trade secret protections bar disclosure of details of their algorithms in the courtroom. The method used to weigh the answers provided to the questionnaire that lies at the heart of the COMPAS predictive model, developed by Northpoint and currently used by many jurisdictions in the U.S., remains confidential.[120] Consequences of this commercial partnership in creating scientific knowledge can exacerbate an observed pattern in disclosure of expert uncertainty in scenarios where experts anticipate being criticized for communicating uncertainty. In the absence of clear assurances that designers of the models will be protected if they disclose uncertainties in the models, itself a condition that is not likely because of the boundary problems discussed before, experts may resort to tools such as protection of trade secrets that undermine communication of uncertainty. Moreover, pervasive attachment to strict trade secrets practice, in conjunction with the movement toward commercializing research and development of these algorithms, largely constricts discourse over residual uncertainty even within the computer science discipline and among experts.[121] Public demand for a market-driven approach to design algorithms will place the most promising research outside of academia and in the private market where developers are siloed in different companies—their work protected by trade-secrets and barred from

115. *See Science, Technology, and Law*, ENCYCLOPEDIA SCI., TECH., & ETHICS (2005), https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/science-technology-and-law (stating that increased technology leads to increased litigation and liability).

116. *See* Carlson, *supra* note 28, at 319–21 (describing application of criminal sentencing algorithms and whether it violates due process).

117. Massachusetts v. Envtl. Prot. Agency, 549 U.S. 497, 516–17 (2007).

118. Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 821 (2014).

119. *See* Carlson, *supra* note 28, at 322–23 (stating defendants are unable to challenge the accuracy of results because the developer considers the algorithm to be a trade secret).

120. *Id.* at 316.

121. *See id.* at 321–22 (describing the use of trade secrets to avoid disclosure of sentencing algorithms).

discourse with other experts in academia.[122] Adding to this significant hurdle in communication of uncertainty, some developers have also argued that lack of communication with regards to uncertainty of current algorithms is due to their expectation that uncertainties will not be understood.[123] Northpoint, one of the biggest companies currently involved in designing algorithms, has responded to criticism of uncertainty embedded in COMPAS models by saying, "[t]here's no secret sauce to what we do; it's just not clearly understood."[124]

Moreover, the problem of black box, unique to the science of algorithms that incorporate machine learning, leads to scenarios where even experts and programmers themselves are not equipped to express their uncertainties.[125] Machine learning models that prove useful are useful precisely because they possess a degree of complexity in analyzing data that evades even the analytical understanding of the engineers and programmers.[126] An unusually intractable communication gap between the experts and the laymen over understanding complexity and uncertainty already exists in the field of computer science because of the specific role played by the language of coding and computers, which is unfamiliar to the general public.[127] Writing for computational devices demands "a special exactness, formality, and completeness that communication via human languages does not," and this renders code language "inaccessible to the majority of the population."[128] Black boxing adds yet another level of opaqueness to this communication challenge and severely exacerbates the problem of communicating uncertainty. This is in turn will further isolate the general public from shaping the dialogue over the use of algorithms.

IV. POLICY PROPOSALS

Heavy reliance on the science-led approach to criminal enforcement based on retribution implicitly forecloses further exploration of alternative approaches to criminal enforcement such as community-based solutions and restorative justice.[129] Unlike the retributivist approach to criminal justice, which focuses on punishment and deterrence—and therefore stands to benefit from development of scientific tools such as algorithms that effectuate those goals—restorative justice finds the most appropriate response to criminal behavior to be repairing the harm caused by the wrongful act.[130] This is done through providing a platform for those directly affected by the crime to come together and discuss the event and avenues for reparation. If the public policy approach to criminal

122. Fischhoff, *supra* note 112, at 7.
123. *Id.* at 4.
124. Carlson, *supra* note 28, at 321.
125. Fischhoff, *supra* note 112, at 7–8.
126. Burrell, *supra* note 62, at 5.
127. *Id.* at 3.
128. *Id.* at 4.
129. *See* Angèle Christin et al., *Courts and Predictive Algorithms*, DATA & C.R.: A NEW ERA OF POLICING & JUST. (Oct. 27, 2015), www.law.nyu.edu/sites/default/files/upload_documents/Angele%Christin.pdf (stating that law enforcement algorithms focus on one theory of justice at the exclusion of others).
130. Jeff Latimer et al., *The Effectiveness of Restorative Justice Practices: A Meta-Analysis*, 85 PRISON J. 127, 128 (2005).

justice utilizes the scientific use of algorithms and models, in conjunction with a healthy balance of restorative justice, victims and offenders can be given a more direct voice in applications of the models in each case.[131] Such an approach is cognizant of another reoccurring theme in science and technology studies, namely that members of the general public have their own knowledge which may complement or rival expert conceptions.[132] This is especially true of the knowledge that could be imparted by those most directly affected by a public policy, or victims in the context of criminal justice reform.[133] The developers and policymakers could be largely blind to the consequences and outcomes that might arise out of a pervasive social-awareness of a machine-centric approach to criminal justice, at the expense of a victim-centric approach. Is a purely science-led approach to crime prevention itself likely to serve as a feedback loop that perpetuates the trends it seeks to prevent, by eroding social ties that would emerge between judges, the community, and the offenders in the alternative system where these parties work together to solve local criminal justice issues? The answer will not be found in the models.[134]

A less ambitious solution would be to shift funding priorities to develop publicly owned risk assessment algorithms, in order to avoid the problems associated with trade secrets.[135] Under this approach, it is conceivable that civic society and community organizers can play a more active role in injecting the community's concerns into what factors should be considered in the models and what weight to be given to each factor, thus mitigating some of the complexities of PUS. Another proposal would ensure that algorithms that incorporate factors such as past criminal behavior also utilize sober and credible assessments about how current societal conditions such as disparities in criminal enforcement may be affecting those factors.[136]

V. CONCLUSION

The best time to construct and shape the ethics of new technologies is before its complexities and uncertain qualities have shaped the interactions between the relevant social groups and have made the system harder to deconstruct.[137] American society should not close its eyes to the inherent human

131. *See id.* (discussing aspects of restorative justice, including the focus on discussion between the victims and offenders).

132. Brian Wynne, *Knowledges in Context*, 16 SCI., TECH., & HUM. VALUES, 111, 111 (1991); Steven Yearley, *Computer Models and the Public's Understanding of Science: A Case-Study Analysis*, 29 SOC. STUD. SCI. 845, 847 (1999).

133. *See* Yearley, *supra* note 132, at 848 (demonstrating that knowledge of those most directly affected by a public policy may complement or rival expert conceptions).

134. *See* Carlson, *supra* note 28, at 321 ("As the demand for risk assessment tools has grown, the private sector has drastically expanded its involvement in designing and marketing risk assessment tools to sell to the government."); Sarewitz, *supra* note 75, at 1001 (observing that feedback loops created through participation of the marketplace in funding and developing scientific research to be applied to social problems repeatedly reduces the role of other social and institutional mechanisms in defining and shaping knowledge).

135. Carlson, *supra* note 28, at 327.

136. Elizabeth Glazer et al., *Debating Risk-Assessment Tools*, MARSHALL PROJECT (Oct. 25, 2017, 2:29 PM), https://www.themarshallproject.org/2017/10/25/debating-risk-assessment-tools.

137. Ananny, *supra* note 67, at 4.

biases that originally incentivized many communities to switch from a human-centric approach to a broader use of models and understand that integration of algorithmic thinking into criminal enforcement decisions can mitigate some of the worst aspects of human bias.[138] Yet, in selecting, filtering, combining, and transforming the input data, algorithms create, sustain, and signify relationships among people and data by using as a computational tool the logic architecture that has been created and shaped by human culture. They are therefore not entirely immune to bias and are indeed designed to reveal and deliver the patterns that preexist in our societies, many of which are undesirable and even widely unknown to the members of the society. Unique problems in communicating the uncertainty of the science of algorithms and potentially unhealthy boundary work implications significantly raises the possibility that the use of algorithms in the courtroom will not be constrained by healthy public input. Public policymakers should therefore approach the decision of integrating machines into the criminal system with a high degree of caution and look at proposals that will increase civic participation in determining the exact place for algorithms in our criminal justice system.

138. Spielkamp, *supra* note 58.