# ALGORITHMIC LEGAL METRICS

*Dan L. Burk**

*Predictive algorithms are increasingly being deployed in a variety of settings to determine legal status. Algorithmic predictions have been used to determine provision of health care and social services, to allocate state resources, and to anticipate criminal behavior or activity. Further applications have been proposed to determine civil and criminal liability or to "personalize" legal default rules. Deployment of such artificial intelligence (AI) systems has properly raised questions of algorithmic bias, fairness, transparency, and due process. But little attention has been paid to the known sociological costs of using predictive algorithms to determine legal status. A large and growing social science literature teaches the effects of "algorithmic living," documenting how humans interact with machine generated assessments. Many of these interactions are socially detrimental, and such corrosive effects are greatly amplified by the increasing speed and ubiquity of digitally automated algorithmic systems.*

*In this Article I link the sociological and legal analysis of AI, highlighting the reflexive social processes that are engaged by algorithmic metrics. This Article examines these overlooked social effects of predictive legal algorithms and contributes to the literature a vital fundamental but missing critique of such analytics. Specifically, this Article shows how the problematic social effects of algorithmic legal metrics extend far beyond the concerns about accuracy that have thus far dominated critiques of such metrics. Second, it demonstrates that corrective governance mechanisms such as enhanced due process or transparency will be inadequate to remedy such corrosive effects, and that some such remedies, such as transparency, may actually exacerbate the worst effects of algorithmic governmentality. Third, the Article shows that the application of algorithmic metrics to legal decisions aggravates the latent tensions between equity and autonomy*



* Chancellor's Professor of Law, University of California, Irvine. My thanks to a patient and insightful legion of careful readers for their helpful comments on earlier iterations of this paper: to Stephanie Bair, Clark Asay, and participants in the January 23, 2020, Brigham Young University Faculty Workshop; to Bob Brauneis, David Fagundes, Andrew Gilden, Glynn Lunney, Zahr Said, Matthew Sag, Christopher Yoo, and participants in the 4th Annual Copyright Roundtable at the University of Pennsylvania; to Roger Ford, Rebecca Wexler, Alison Cooper, and participants in the April 6, 2019, Yale ISP conference on "(Im)Perfect Enforcement"; to Swethaa Balakrishnen, Rick Hasen, Vic Fleischer, Emily Taylor Poppe, Tony Reese, Ken Simons, Shauhin Talesh, and participants in the January 30, 2019, UCI Law Faculty Works in Progress workshop; to Micheal Birnhack, Assaf Jacob, and participants in the January 6, 2019, TAU Law & Technology Workshop; and to participants at the 9th Annual ICIL conference at University of Antwerp. Generous research support was provided by a grant from the The Honorable Nancy F. Atlas Intellectual Property American Inn of Court, administered by the University of Houston Institute for Intellectual Property and Information Law.

*in liberal institutions, undermining democratic values in a manner and on a scale not previously experienced by human societies. Illuminating these effects casts new light on the inherent social costs of AI metrics, particularly the perverse effects of deploying algorithms in legal systems.*



## INTRODUCTION

Automated pattern analysis and decisionmaking, colloquially designated as "artificial intelligence" or "AI," is increasingly being deployed to mediate or to assist in social determinations across a range of domains including governance and regulatory decisions.[1] As potential applications for algorithmic legal decisionmaking grow, optimistic visions of such systems foresee the rise of accurate and efficient AI regulators, free from the errors of human decisionmakers.[2] More pessimistic visions foresee the imposition of impersonal and regimented machine discipline on an unsuspecting populace.[3]

---

1 *See* Adrian Mackenzie, *The Production of Prediction: What Does Machine Learning Want?*, 18 EUR. J. CULTURAL STUD. 429, 430 (2015); Monika Zalnieriute, Lyria Bennett Moses & George Williams, *The Rule of Law and Automation of Government Decision-Making*, 82 MOD. L. REV. 425, 427–28 (2019).

2 *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1155 (2017) [hereinafter Coglianese & Lehr, *Regulating by Robot*] (arguing that governmental reliance on machine learning should be approached with measured optimism); Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 54–55 (2019) [hereinafter Coglianese & Lehr, *Transparency*] (arguing that algorithmic governance can be transparent, accurate, and efficient); Zalnieriute et al., *supra* note 1, at 454 (summarizing the potential benefits of automated decisionmaking); *see also* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 30–32 (2013) (touting the supposed accuracy and comprehensiveness of "Big Data" analyses).

3 *See, e.g.*, VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2017); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); *see also* Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, 12

Despite the confluence of such algorithmic hope and dread, both public and private legal functions are increasingly the subjects for algorithmic provision.[4] Predictive algorithms have been deployed to identify families at risk of abusive behavior, in order to mobilize social services intervention before actual harm occurs.[5] Predictive algorithms have been relied upon to assess the threat of criminal recidivism, and so determine the allowance for bail or for prisoner parole.[6] Predictive algorithms are being incorporated into policing strategies, allowing law enforcement resources to be positioned where criminal activity is anticipated to occur.[7] And algorithmic predictions are becoming progressively arrayed across a broad swath of other legal and social decisionmaking: to allocate public assistance,[8] to preempt customs and border violations,[9] to determine immigration status,[10] to forecast threats to national security.[11]

Emerging proposals suggest an even greater role for algorithmically determined legal metrics. The confluence of massively multisourced consumer surveillance and machine learning technologies has led to proposals for algorithmically mediated "personalized law" in a variety of public and private law areas.[12] Specifically, recent scholarship has suggested that the collection of detailed information on consumers, together with algorithmic processing of such data, will allow for customized tailoring of legal impera-

---

REGUL. & GOVERNANCE 505, 513 (2018) (observing that "[a]lgorithmic systems have . . . been associated with two dramatically opposed political visions").

4 *See* Coglianese & Lehr, *Regulating by Robot*, *supra* note 2, at 1151; Lyria Bennett Moses, *Artificial Intelligence in the Courts, Legal Academia and Legal Practice*, 91 AUSTRALIAN L.J. 561 (2017).

5 *See* LINA DENCIK, ARNE HINTZ, JOANNA REDDEN & HARRY WARNE, DATA JUST. LAB, DATA SCORES AS GOVERNANCE: INVESTIGATING USES OF CITIZEN SCORING IN PUBLIC SERVICES: PROJECT REPORT 11, 27, 55 (2018); EUBANKS, *supra* note 3, at 132, 140–47.

6 *See* State v. Loomis, 2016 WI 68, ¶¶ 10–21, 881 N.W.2d 749, 754–55 (evaluating use of an algorithm to predict criminal recidivism); Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (discussing use of algorithms to predict criminal recidivism).

7 *See* DENCIK ET AL., *supra* note 5, at 74; Lyria Bennett Moses & Janet Chan, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, 28 POLICING & SOC'Y 806, 813 (2018).

8 *See* DENCIK ET AL., *supra* note 5, at 48, 52; EUBANKS, *supra* note 3, at 76–82.

9 FAIZA PATEL, RACHEL LEVINSON-WALDMAN, SOPHIA DENUYL & RAYA KOREH, BRENNAN CTR. FOR JUST., SOCIAL MEDIA MONITORING: HOW THE DEPARTMENT OF HOMELAND SECURITY USES DIGITAL DATA IN THE NAME OF NATIONAL SECURITY 15–16 (2019).

10 *Id.* at 27–29.

11 *Id.* at 15, 20–21.

12 *See generally* Anthony J. Casey & Anthony Niblett, *A Framework for the New Personalization of Law*, 86 U. CHI. L. REV. 333 (2019) (surveying the literature on algorithmically personalized law); Philipp Hacker, *Personalizing EU Private Law: From Disclosures to Nudges and Mandates*, 25 EUR. REV. PRIV. L. 651 (2017) (surveying personalized law literature under European legal systems).

tives to the capacity or characteristics of individual actors.[13] This body of work argues that legal directives could be matched to detailed consumer profiles so as to create metrics that are "personalized" for the profiled individual, rather than uniform for the general populace.[14] Rather than evaluating the standard of care for a hypothetical reasonably prudent person, tort law could perhaps algorithmically determine the standard of care for a given accused tortfeasor.[15] Rather than allocate inheritance according to default intestacy rules, estate law could perhaps devise assets according to the algorithmically predicted preferences of a given decedent.[16] Proposals of this sort have been circulated for a variety of other legal regimes, including contract, criminal law, and copyright.[17]

Relying as they do on mechanisms of consumer surveillance, these proposals are effectively intended to translate the growing provision of mass "personalization" or "modulation" of market services and institutions to the provision of legal services and institutions.[18] Although such proposals for personalized legal metrics carry a degree of superficial plausibility, on closer inspection it becomes clear that they combine the worst defects of idealized economic analysis and simplistic algorithmic utopianism.[19] Such proposals display a breathtaking degree of naiveté regarding the workings of algorithmic classification, not merely regarding the limitations of the techni-

---

13 *See* Omri Ben-Shahar & Ariel Porat, *Personalizing Negligence Law*, 91 N.Y.U. L. REV. 627, 674–75 (2016) (postulating that personalized negligence standards could be generated via predictive algorithms); Ariel Porat & Lior Jacob Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417, 1433–35, 1442–44 (2014) (postulating that personalized legal rules for wills, organ donation, and other matters could be generated via predictive algorithms).

14 *See, e.g.*, Anthony J. Casey & Anthony Niblett, *Self-Driving Laws*, 66 U. TORONTO L.J. 429 (2016) [hereinafter Casey & Niblett, *Self-Driving Laws*]; Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401 (2017).

15 *See* Ben-Shahar & Porat, *supra* note 13, at 674–75.

16 *See* Porat & Strahilevitz, *supra* note 13, at 1419–20.

17 *See, e.g.*, *id.* at 680; Omri Ben-Shahar & Ariel Porat, *Personalizing Mandatory Rules in Contract Law*, 86 U. CHI. L. REV. 255, 256 (2019); Matthew B. Kugler & Lior Jacob Strahilevitz, *Assessing the Empirical Upside of Personalized Criminal Procedure*, 86 U. CHI. L. REV. 489, 490–91 (2019); Adi Libson & Gideon Parchomovsky, *Toward the Personalization of Copyright Law*, 86 U. CHI. L. REV. 527, 528 (2019); Porat & Strahilevitz, *supra* note 13, at 1419.

18 *See* Julie E. Cohen, *Between Truth and Power*, *in* INFORMATION, FREEDOM AND PROPERTY: THE PHILOSOPHY OF LAW MEETS THE PHILOSOPHY OF TECHNOLOGY 57, 62, 66 (Mireille Hildebrandt & Bibi van den Berg eds., 2016) (explaining the market strategy of modulated personalization to gain competitive advantage); Karen Yeung, *Five Fears About Mass Predictive Personalization in an Age of Surveillance Capitalism*, 8 INT'L DATA PRIV. L. 258, 259–60 (2018) (explaining the industrial movement from mass production to mass personalization).

19 *Cf.* Philip Maximilian Bender, Limits on Personalization: A Normative Theory of Default Rule Tailoring 8–9 (Apr. 2019) (unpublished manuscript) (on file with author) (arguing that such personalized law proposals fail both normative legal and constitutional requirements).

cal infrastructure on which such classifications would rest,[20] but regarding the characteristics of the social infrastructure on which such classifications depend. An increasingly robust sociological literature demonstrates that algorithmic scoring effectuates "classification situations" that recreate and reinforce existing social orders, accelerating some of the most problematic mechanisms for exploitation and inequality.[21] Such metrics not only amplify and reinforce existing social biases, but tend to produce detrimental self-surveillance. Due to such effects, the quantified assessments supplied by algorithmic scoring are not neutral, but take on normative and moral connotations.

As yet, the legal policy discussions on algorithmic decisionmaking have taken little note of this work. But given the growing literature demonstrating the perverse social effects of algorithmic scoring systems, it seems clear that the incorporation of such metrics into the determination of legal status offers a new and troubling challenge to the rule of law. Legal determinations such as tort liability or criminal culpability that carry their own moral weight are likely to produce unintended consequences when associated with morally charged algorithmic metrics. A close examination of these mechanisms quickly illuminates disjunctions at the intersection among jurisprudence, automated technologies, and socially reflexive practices, and alerts us to areas of concern as legal institutions are increasingly amalgamated into the growing algorithmic assemblage.

The existing legal literature has only begun to touch the most evident issues regarding algorithmic governance. Following the categorical framework laid out by Lucas Introna, we might divide the issues of governance, and the legal literature addressing such issues, into three groupings.[22] The first of these categories concerns *governance by algorithms,* that is, the effects of deploying automated systems to administer legal and regulatory oversight.[23] Related to this set of questions, we can discern another emerging literature addressing a second set of issues around *governance of algorithms*, that is, the problems and issues related to oversight of algorithms that are deployed in both public and private sectors.[24] Under the first set of questions, we want to investigate whether automated systems used in governance are likely to promote efficiency, justice, equity, and democratic values. Under the second set of questions, we want to consider how to ensure that the operation of automated systems is fair, accurate, unbiased, and legally compliant. The inquir-

20 *See generally* A. Feder Cooper & Karen E.C. Levy, Imperfection Is the Norm: A Computer Systems Perspective on Real-Time IoT and Enforcement (Apr. 2019) (unpublished manuscript) (on file with author) (discussing the fundamental engineering limitations on latency and accuracy in distributed computing).

21 *See* Angwin et al., *supra* note 6; *see also infra* notes 148–53 and accompanying text.

22 *See* Lucas D. Introna, *Algorithms, Governance, and Governmentality: On Governing Academic Writing,* 41 SCI. TECH. & HUM. VALUES 17, 28–30 (2016).

23 *Id.* at 29.

24 *Id.* at 28–29.

ies are clearly related, as for example, in the concern that inaccurate or biased algorithms are unlikely to produce fair or just social outcomes.

Each of these sets of inquiries constitutes a legitimate and important line of investigation, but neither is my primary concern in this Article. Instead, I focus here on a third set of issues that has gone virtually unaddressed in the legal literature. Borrowing a term from Foucault, we may term these to be questions relating to the *governmentality of algorithms*, that is, to the mechanisms by which algorithms may fundamentally alter the personal behaviors and social structures with which they interact.[25] Under the first two sets of inquiries, previous commentators have begun to consider whether algorithmic governance comports with the rules and expectations we have for a civil society. But here I hope to address the antecedent question as to when the deployment of algorithmic systems may fundamentally change the rules and expectations by which we live. The question is not whether algorithms can or do fall within the rules; the question is how and whether they *make* the rules.

Consequently, in this Article, I begin to map out the intersection between the social effects of quantification and the social construction of algorithms in the context of legal decisionmaking. In previous work, I have explored the implications of attempting to incorporate legal standards into algorithms, arguing that the social action typical of algorithmic systems promises to shape and eventually become the legal standard it seeks to implement.[26] Here I essentially consider the inverse proposition: I explore the effects of incorporating algorithms, which is to say algorithmic metrics, into legal standards. In particular, I will examine the anticipated use of algorithmically processed "Big Data" in attempting to align legal incentives with social expectations.

I begin my examination by sketching the features of the sprawling and uncoordinated data gathering apparatus, or "surveillant assemblage," from which the profiles for algorithmic processing are extracted. I particularly highlight the distortions introduced into data profiles by the processing, by the social context, and by the inevitable interpretation associated with algorithmic metrics. A number of previous commentators have been properly concerned about the biases endemic to data profiling, but I argue that algorithmic bias goes well beyond the previous discussions of prejudice or inaccuracy to shape and define the social relationships and behavior surrounding the subjects of algorithmic data profiling.

Beginning in Part II, I locate the source of such distortions in reflexive social practices that are embodied in algorithmic measurements, and with which algorithmic processes interact in a broader structural context. Although there are numerous case studies documenting such effects, I employ familiar illustrations drawn from the intensively studied examples of commensurate credit scoring and law school ranking. The reflexive effects

25 *Id.* at 30.

26 *See* Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI. L. REV. 283, 285 (2019).

found in such algorithmic processes are well known in the existing social science literature, but are now accelerated and amplified by the speed and scale of automated data analysis and processing. In Part III, I argue that due to such processes, algorithmic metrics are performative, in the sense that they create and instantiate the numerical models on which they are based. In other words, algorithms create their own social facts. These effects are particularly acute where algorithms instantiate economic constructs, where they can be seen to reconfigure the participants and the practices in modern commercial exchanges. I show that such effects are, perhaps paradoxically, heightened by transparency of algorithmic inputs and processes, leaving in doubt the advisability of some recent scholarly calls for greater transparency in algorithmic profiling.

Finally, in the last two Parts of the Article, I link these concepts to the normative functions of law, showing how legal judgments will be distorted by the introduction of algorithmic scoring regimes, particularly those being imported from datafied business models in the private sector. I describe how the social processes on which algorithmic metrics rest lead ultimately to the characterization of such metrics as moral character judgments. When inserted into legal determinations that intrinsically require moral character judgments, we may expect the value biases embedded in algorithmic legal metrics to effectively become legal judgments. The precipitation of algorithmic metrics into legal culpability poses a particular problem for American legal discourse, due to the American legal system's decades-long fascination with the economic analysis of law.

In tracing the characteristic arc of algorithmic metrics from profiling through legal application, this Article makes several novel contributions to the literature on law and algorithmic governance. First, it details corrosive social effects of algorithmic legal metrics that extend far beyond the concerns about accuracy that have thus far dominated critiques of such metrics. Second, it demonstrates that deploying traditional corrective governance mechanisms such as enhanced due process or transparency is wholly inadequate to remedy such corrosive effects, and indeed that some such remedies, such as transparency, may actually serve to exacerbate the worst effects of algorithmic governmentality. Third, the Article shows that the application of algorithmic metrics to legal decisions aggravates the latent tension between equity and autonomy that is endemic in liberal institutions, undermining democratic values in a manner and on a scale not previously experienced by human societies. I therefore close with some thoughts regarding restriction of algorithmic quantification in legal settings, and the need to identify the areas most perniciously affected by such systems, so as to exclude or curtail automated decisionmaking from such decisions.

## I. Surveillance and Profiling

In one sense, algorithmic regulatory practice has deep historical roots, and so might be thought simply to constitute an extension of well-known governance practices embedded in the modern bureaucratic state.[27] Beginning with the development of nineteenth-century bureaucratic governance, statistical methods have been increasingly used to assess and manage populations, defining the existence of persons according to actuarial criteria.[28] Nineteenth-century data was gathered with some statistical notion as to their purpose—assessing the incidence of crime, or of disease, or of financial transactions.[29] More modern data accumulation and processing practices continue to gather all these statistical records and more, but they are repurposed toward the management of populations through statistical and demographic categorization that Foucault termed "biopower."[30] Thus, current practices mark a shift from quantification of social statistics in order to *describe* and *predict* relationships to quantification of social relationships in order to *monitor* and *control* them.[31]

This shift in the purpose of social quantification has been dramatically advanced by the technical capabilities of recordkeeping media and methods. Once either public or private entities began amassing records, it was not long before the accumulated data contained in paper records could be recombined for new purposes not anticipated in their compilation.[32] Such recombination might occur within an organization, then between organizations, and then between organizations with very different goals and provenance, such as public governmental and private corporate institutions.[33] The rate and frequency of recombination is largely a function of the medium, which with physical written records is relatively slow and intermittent. But the potential for recombination becomes exponentially greater with the speed

27 *See* THEODORE M. PORTER, TRUST IN NUMBERS: THE PURSUIT OF OBJECTIVITY IN SCIENCE AND PUBLIC LIFE 7–8 (1995) (discussing the reliance of unelected bureaucrats on ostensibly objective numerical criteria in order to gain decisional legitimacy); *see also* Luciana Parisi, *Critical Computation: Digital Automata and General Artificial Thinking*, 36 THEORY CULTURE & SOC'Y, no. 2, 2019, at 89, 92 (discussing the relation of modern machine learning techniques to the history of statistical pattern recognition).

28 *See* Ian Hacking, *Biopower and the Avalanche of Printed Numbers*, 5 HUMANS. SOC'Y 279 (1982); Ian Hacking, *Making Up People*, *in* RECONSTRUCTING INDIVIDUALISM: AUTONOMY, INDIVIDUALITY, AND THE SELF IN WESTERN THOUGHT 222, 223 (Thomas C. Heller, Morton Sosna & David E. Wellbery eds., 1986) [hereinafter Hacking, *Making Up People*]; Peter Miller & Nikolas Rose, *Governing Economic Life*, 19 ECON. & SOC'Y 1, 12 (1990).

29 *See* LORRAINE DASTON, CLASSICAL PROBABILITY IN THE ENLIGHTENMENT, at xii (1988); IAN HACKING, THE TAMING OF CHANCE 3 (1990).

30 MICHEL FOUCAULT, "SOCIETY MUST BE DEFENDED": LECTURES AT THE COLLÈGE DE FRANCE, 1975–76, at 243, 250, 254 (Mauro Bertani & Alessandro Fontana eds., David Macey trans., 2003).

31 *See* MICHEL FOUCAULT, DISCIPLINE & PUNISH: THE BIRTH OF THE PRISON 183–84 (Alan Sheridan trans., 1977); *see also* Mackenzie, *supra* note 1, at 434 (noting that modern machine learning is reliant on the extension of methods developed in the nineteenth century).

32 Stanton Wheeler, *Problems and Issues in Record-Keeping*, *in* ON RECORD: FILES AND DOSSIERS IN AMERICAN LIFE 3, 5 (Stanton Wheeler ed., 1969).

33 Kevin D. Haggerty & Richard V. Ericson, *The Surveillant Assemblage*, 51 BRITISH J. SOCIOLOGY 605, 610–11 (2000).

and enormous storage capacity of digitized records. As data recombination becomes easier, it becomes almost an end in itself, with justifications generated after the collection, dissemination, combination and correlation has already occurred.

The literature on the implications of such electronic data profiling is enormous, and a full exploration lies well beyond the scope of my discussion in this Article.[34] Here I wish to highlight only a few salient points that are important to the present investigation. First, algorithmic profiling from recombinant data records relies upon the increasingly intensified forms of technological observation operated by both state and private entities.[35] Individual behavior is profiled from combinations of police records, court dockets, insurance profiles, shopping behavior, credit scoring, geolocation patterns, vehicle and land title registrations, vital statistics, electronic communications, browsing habits, search queries, and multiple other routine social transaction records.[36] In some cases data is combed from the digitized records of physical transactions; in other cases it is extracted from the stored digital traces of purely electronic occurrences. Digital formats make records of each type easier to combine with records of similar origin, or of different origins.

Second, the concept of comprehensive, "panoptic" surveillance has become the standard trope in discussing the burgeoning incidence of data accumulation, storage, and analysis.[37] The reference has its origin in Bentham's famous concept of the "panopticon," envisioned as a surveillance system for prisons, intended to impose behavioral conformity among prisoners by the effect of observational uncertainty: because the prisoners might be surreptitiously observed at any time, they would behave as if they were observed at all times.[38] This form of self-monitoring was later recognized by Foucault as a type of institutional power to impose order by means of internalized discipline—not only prisoners, but citizens in the general population could be induced to change their own behavior in response to surveillance.[39] In essence, Bentham's original, particular penal setting for panoptic control has become inconsequential, as a population under continuous surveillance becomes prisoner of its own responses.

34 For a few major guideposts, see, for example, JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 4 (2012); HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 6 (2010); DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 2 (2004).

35 *See* Haggerty & Ericson, *supra* note 33, at 610.

36 *Id.* at 613, 616–18.

37 *See, e.g.*, Jerome E. Dobson & Peter F. Fisher, *The Panopticon's Changing Geography*, 97 GEOGRAPHICAL REV. 307, 309 (2007); Joshua Fairfield, *Escape into the Panopticon: Virtual Worlds and the Surveillance Society*, 118 YALE L.J. POCKET PART 131, 131–32 (2009); Elizabeth Stoycheff, Juan Liu, Kai Xu & Kunto Wibowo, *Privacy and the Panopticon: Online Mass Surveillance's Deterrence and Chilling Effects*, 21 NEW MEDIA & SOC'Y 602, 604 (2019).

38 *See* Haggerty & Ericson, *supra* note 33, at 607.

39 *See* FOUCAULT, *supra* note 30, at 195–97.

Third, these disciplinary effects are enormously multiplied by the availability of automated data processing. As the speed and capacity of computational processing is added to surveillance practices, the general populace is increasingly the subject of the widespread, computer-enabled collection, searching, and screening of digital records that Gandy famously dubbed the "panoptic sort."[40] Surveillance, sorting, and processing capabilities reinforce and feed on one another. The torrent of available data can only be processed by superhuman, automated means, and in turn the availability of such automated systems invites the continued collection of surveillant data.

Fourth, the vast quantities of accumulating surveillant data are no longer necessarily sorted according to criteria contemplated by human controllers, but rather according to criteria detected and determined by machines. These systems are not "artificially intelligent" in any robust meaning of that term.[41] Because of intractable problems in directly coding judgment and reasoning processes, attention in AI design has increasingly turned to machine learning systems, by which statistical algorithms discern patterns in massive datasets, and optimize their functions according to those patterns.[42] When used for predictive analytics, such systems identify category characteristics in order to similarly classify future data with congruent characteristics.[43] But rather than having a programmer specify a categorical model, the machine develops its own model via progressive rounds of rapid trial and error.[44] One result of such endogenous machine categorization is that the criteria for the outcome often remain unexplained, and are often indecipherable to human explication.[45]

Fifth, the vast and growing apparatus of surveillance is the result of the natural combination and intersection of ostensibly independent electronic monitoring and recordkeeping.[46] Independently created systems are joined and connected out of convenience, efficiency, or necessity. Concatenated monitoring and recordkeeping apparatus arise from the intersection of

---

40 *See* OSCAR H. GANDY, JR., THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION 15 (1993).

41 *See* M.C. Elish & danah boyd, *Situating Methods in the Magic of Big Data and AI*, 85 COMMC'N MONOGRAPHS 57, 61–63 (2018) (describing the current emphasis on machine learning over "good old fashioned" artificial intelligence (quoting JOHN HAUGELAND, ARTIFICIAL INTELLIGENCE: THE VERY IDEA 112 (1985))); *see also* Marion Fourcade & Kieran Healy, *Seeing Like a Market*, 15 SOCIO-ECON. REV. 9, 24 (2017) (observing that AI research abandoned the idea of machines that can think in favor of machines that can learn).

42 Marion Fourcade & Kieran Healy, *Categories All the Way Down*, 42 HIST. SOC. RSCH. 286, 292–93 (2017).

43 *See* Mackenzie, *supra* note 1, at 433; *see also* Geoffrey C. Bowker, *The Theory/Data Thing*, 8 INT'L J. COMMUNICATION 1795, 1796–97 (2014) (observing that even though categories are not specified, "Big Data" correlation inevitably implicates categorization).

44 Parisi, *supra* note 27, at 100.

45 Fourcade & Healy, *supra* note 42, at 293; *see also* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1040 (2017) (reviewing PASQUALE, *supra* note 3) ("[I]n the era of self-enhancing algorithms, the algorithm's human designers may not fully understand their own creation . . . .").

46 Haggerty & Ericson, *supra* note 33, at 610–11.

diverse systems, each converging into the next, lacking discernable boundaries. Thus, no single technology or social system constitutes the origin or genesis of panoptic surveillance.[47] And by the same token, the growing concern over privacy and surveillance cannot be easily answered or resolved by curtailing or modulating a given practice, because privacy erosion cannot be attributed to any one source.[48]

Instead, a proliferating conglomeration of previously discrete and separate surveillance systems functions as what has been termed in total the "surveillant assemblage."[49] The assemblage designation recognizes that beneath the seemingly stable exterior of any entity lies a complex tangle of multiple phenomena and processes working in concert. An assemblage thus constitutes a conglomeration of myriad interconnected tangible and intangible components that function together despite their heterogeneous origins.[50] With the addition of AI analysis and processing for surveillance data from disparate sources, the resulting sociotechnical sprawl might be designated the "algorithmic assemblage," and our goal here is to discern the circumstances of legal status assigned within its ambit.[51]

### A. *Big Data Processing*

A key consideration in the deployment of algorithmic legal metrics must be the assumptions built into their operation. As this Section discusses in detail, the machine learning systems on which such metrics are based are by no means neutral, omniscient, or infallible. They are to the contrary highly selective, incomplete, and deeply value laden. Their technical design models the world under assumptions that may be profoundly at odds with their use in legal determinations.

Unfortunately, some commentators observing the increasing availability of massive datasets have mistakenly supposed that algorithmic analysis of such copious information offers exhaustive or comprehensive results.[52] This

---

47 *Id.* at 610.

48 *Id.* at 609.

49 *Id.* at 606.

50 Manuel DeLanda, A New Philosophy of Society: Assemblage Theory and Social Complexity 10–12 (2006); Gilles Deleuze & Félix Guattari, A Thousand Plateaus: Capitalism and Schizophrenia 71, 88–91 (Brian Massumi trans., 1987).

51 *Cf.* N. Katherine Hayles, Unthought: The Power of the Cognitive Nonconscious 118 (2017) (describing as a "cognitive assemblage" the "arrangement of systems, subsystems, and individual actors through which information flows" between human and algorithmic actors).

52 *See, e.g.*, Mayer-Schönberger & Cukier, *supra* note 2, at 33–34; *see also* Casey & Niblett, *Self-Driving Laws*, *supra* note 14, at 437 (asserting that "[t]he biases and inconsistencies found in individual judgments can largely be washed away using advanced data analytics"). *Contra* Carl Lagoze, *Big Data, Data Integrity, and the Fracturing of the Control Zone*, Big Data & Soc'y, July–Dec. 2014, at 1, 5 (critiquing the claim by Mayer-Schönberger and Cukier that massive datasets offer statistically complete and faultless outputs); S. Leonelli, *What Difference Does Quantity Make? On the Epistemology of Big Data in Biology*, Big Data & Soc'y, Apr.–June 2014, at 1, 2 (same).

is a fundamental misconception. Data analysis is necessarily selective, not promiscuous. Certainly, it is true that the evolving surveillant assemblage feeds on a wide variety of inputs: text, video, audio, statistical, transactional, positional, and other information varieties.[53] But each of these diverse sources entails its own particular structural and content formats. The use of such disparate data sources in profiling requires the imposition of some common configuration on differently structured or unstructured digital data. And imposition of a common standard on heterologous datasets inevitably results in reconfiguration of the data, and discarding of a good deal of information.

Thus, as Amoore and Piotukh observe, analytic inputs are not collected; the input process is rather properly designated as *ingestion*, with all the biological implications of mastication, decomposition, and metabolism that the term entails.[54] Because data are drawn from a wide variety of sources in numerous formats, analysis of the resulting numerical mélange is possible only by extraction and reformatting of the digital patchwork. Algorithmic data processing employs the spatial logics of the mathematical and physical sciences to reduce a wide range of inputs to homogenous numerical inputs.[55] The conversion of files to standardized, compatible formats reduces differences in *kind* to differences in *degree*.[56]

Data collection and preparation thus require the linkage of data elements by means of *joins* across the intersections and correlations from different datasets. Such linkage and compilation necessarily strip away much of the unique formatting and context of the original source. To reconfigure otherwise incompatible data, analytical processing imposes a radical decontextualization on the data, paring away extraneous information and meanings.[57] Contextual, indexical, symbolic, or experiential differences deemed inessential to the data are subsumed in processing.[58] For example, in text files, words and punctuation that are considered superfluous to the analytical purpose are removed; so too in other types of files, ostensibly irrelevant pixels and bits are distilled off. This process results in a flattening of the ultimate dataset that is indifferent to the qualitative differences contained in the initial data.[59]

---

53 *See* Mackenzie, *supra* note 1, at 433.

54 Louise Amoore & Volha Piotukh, *Life Beyond Big Data: Governing with Little Analytics*, 44 ECON. & SOC'Y 341, 345 (2015).

55 *See* LUCIANA PARISI, CONTAGIOUS ARCHITECTURE: COMPUTATION, AESTHETICS, AND SPACE 3, 8–9 (2013).

56 Amoore & Piotukh, *supra* note 54, at 361.

57 Antoinette Rouvroy, *Technology, Virtuality and Utopia: Governmentality in an Age of Autonomic Computing*, *in* LAW, HUMAN AGENCY AND AUTONOMIC COMPUTING: THE PHILOSOPHY OF LAW MEETS THE PHILOSOPHY OF TECHNOLOGY 119, 126 (Mireille Hildebrandt & Antoinette Rouvroy eds., 2011) (explaining that algorithmic correlation of captured data is indifferent to the causes of phenomena detected).

58 Mackenzie, *supra* note 1, at 434.

59 Amoore & Piotukh, *supra* note 54, at 348–49.

Additionally, as Andrew Iliadis relates, formalized ontologies are often employed to integrate separate datasets, by providing a common categorical rubric for data that is drawn from heterologous formatting and classification schemes. Imposition of a common ontology homologizes the data for comparison and analysis. But by imposing common classification on differing informational formats, new factors are introduced into the resulting dataset. Ontological templates entail their own implicit biases and assumptions, overriding the original dataset with standardized epistemic imperatives.[60]

Ingested files are partitioned and sorted, first broken down into manageable pieces, and then reassembled according to summary inputs drawn from the fragments. A common technique is to shard data files, arbitrarily breaking them up according to size boundaries that do not correspond to their initial structure, then distributing the shards across multiple processors for analysis.[61] Analysis proceeds according to inductively generated queries that look for strong correlations that emerge from the data.[62] Typically, predictive systems assume that there is an underlying pattern in the data that can be described by a mathematical function; the system then begins to generate and test functions to fit the data inputs, seeking useful approximations of the function that is assumed to be there.[63] The algorithm generates different search criteria, selects among them according to relational strength, and makes some determination as to which results are thought to be sufficiently robust or interesting to be elevated as search criteria.[64]

This iterative process has been dubbed "the automation of automation" because the system generates its own search criteria.[65] Rather than predetermined search criteria that are designated via formal symbolic representation of ideas or explicit equations, the analytical algorithms generate iterative search rules and retain those that have the strongest levels of relational support or confidence.[66] It is in some sense "the inverse of programming": the system is not engaged to derive output from a given algorithm, but rather to derive the algorithm that produces a given output.[67] Not all patterns are identified or measured, neither are all patterns preserved for analysis. Rather, the system reduces choices by selecting the most plausible data correlations.[68]

---

60 Andrew Iliadis, *Algorithms, Ontology, and Social Progress*, 14 GLOB. MEDIA & COMMC'N 219, 223 (2018).

61 Ilias Tachmazidis, Grigoris Antoniou, Giorgos Flouris & Spyros Kotoulas, *Scalable Nonmonotonic Reasoning over RDF Data Using MapReduce*, *in* JOINT WORKSHOP ON SCALABLE AND HIGH-PERFORMANCE SEMANTIC WEB SYSTEMS 75, 77 (2012).

62 FRANK OHLHORST, BIG DATA ANALYTICS: TURNING BIG DATA INTO BIG MONEY 8 (2013).

63 Mackenzie, *supra* note 1, at 435.

64 *See* Amoore & Piotukh, *supra* note 54, at 353–54; Elish & boyd, *supra* note 41, at 70–71.

65 Parisi, *supra* note 27, at 90.

66 *Id.*

67 *Id.* at 92.

68 *Id.*

Thus, algorithmic pattern detection and scoring outputs are not found, they are actually *constructed* by the processes of data harvesting, ingestion, and analysis. The process of machine learning develops and connects recurrent data features to indeterminate external factors.[69] This occurs not as a search to discover or identify information of interest; rather, the object of interest is essentially created by the process of analysis. The analytical process detaches the object of interest from the population screened. Statistical identification of relationships determines which patterns in the partitioned and reassembled data become perceptible.[70] The result is to focus attention on whatever the algorithm determines to be an item of interest, while discarding and negating the context from which the object of interest is drawn.[71]

When applied to the data associated with individuals, this process in effect creates identities.[72] Far from representing or capturing any given individual, the action of algorithmic processing results in the generation of abstracted and decorporealized "data doubles" that figure or signify individual actions.[73] This representational digital doppelganger is entirely epiphenomenal, capturing only the record of manifest personal activities.[74] The construct in some sense constitutes a virtual profile,[75] decontextualized and extracted from an amalgam of processed data.[76] It is neither accurate nor inaccurate with regard to any corresponding natural person.[77] It is rather an ex post emblem of assignment of the related natural person to an algorithmically determined category.[78]

For any given data double, the algorithmically created category assigned to it—for credit worthiness, security risk, willingness to pay, or other personal characteristic—may possibly correspond to what is meant by the same term outside of a numerical correlation. But machine learning is based upon cor-

---

69 *Id.* at 99.

70 Amoore & Piotukh, *supra* note 54, at 360.

71 *Id.*

72 John Cheney-Lippold, *A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control*, THEORY, CULTURE & SOC'Y, Nov. 2011, at 164, 165.

73 Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 173–74 (Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot eds., 2014); Haggerty & Ericson, *supra* note 33, at 611, 612.

74 Gillespie, *supra* note 73, at 174.

75 MARK POSTER, THE MODE OF INFORMATION: POSTSTRUCTURALISM AND SOCIAL CONTEXT 97 (1990).

76 David Lyon, *Surveillance as Social Sorting: Computer Codes and Mobile Bodies*, *in* SURVEILLANCE AS SOCIAL SORTING: PRIVACY, RISK, AND DIGITAL DISCRIMINATION 13, 22 (David Lyon ed., 2003); Yeung, *supra* note 3, at 515.

77 Haggerty & Ericson, *supra* note 33, at 614; *see also* Celia Lury & Sophie Day, *Algorithmic Personalization as a Mode of Individuation*, THEORY, CULTURE & SOC'Y, Mar. 2019, at 17, 24 (explaining that predictive algorithms construct iterative approximations rather than establish relations of absolute equivalence).

78 *See* Lury & Day, *supra* note 77, at 20.

relative patterns, and indifferent to the causes of such correlation.[79] Whatever the corresponding categories may mean in physical society, algorithmic processing configures them as mathematical vectors, associating certain data patterns with the statistical doubles generated in the wake of individual behaviors. In this sense, algorithmic categorization defines rather than discovers the parameters of a given population, how that population is discursively situated and perceived, and hence what opportunities or life chances that population will encounter.[80]

### *B. Algorithmic Biases*

It should be clear from the exposition thus far that data analytics are technologies that entail a particular type of perception; just as the natural human organs of perception such as the eye or ear are selectively attuned to particular information, screening out some inputs and reformatting others for comprehension, so the algorithms that make data manageable reduce and reassemble heterogeneous inputs to numerical homogeneity for quantitative processing.[81] In doing so, these analytic processes alter the value and meaning of the data inputs they quantify, and recursively impose those values and meaning on the social processes that rely upon them.[82]

These characteristics of data analytics raise the issue of bias in algorithmic metrics. Because the operations and output of algorithmic processing are expressed quantitively, there is a temptation to think of them as objectively free from biases.[83] This is a common foible with any numerical operation and representation; bare numbers always seem to be free from ideological taint. Moreover, because pattern assessment and targeting are ceded to the algorithm, the process offers additional appearances of neutrality. Because there is no proximate human animus in the determination or execution of the algorithm's determination, the data inputs seem cognitively remote from the level of human fallibility—what Donna Haraway famously referred to as "the god trick."[84] Scored data thus impart an illusion of objectivity because they are rendered by machines, and their human initiators appear distantly removed from the ultimate output.[85]

But far from offering an objective measurement, any system of quantitative accounting necessarily entails its own particular social and political con-

79 *See* Rouvroy, *supra* note 57, at 126; *see also* Bowker, *supra* note 43, at 1795 (observing that algorithmic data methods essentially bypass the maxim that correlation is not causation).

80 Cheney-Lippold, *supra* note 72, at 175.

81 Amoore & Piotukh, *supra* note 54, at 344.

82 *Id.*

83 *See* Rouvroy, *supra* note 57, at 127; *see also* PASQUALE, *supra* note 3, at 15 (observing that algorithmic systems give decisions a "patina of inevitability").

84 Donna Haraway, *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, 14 FEMINIST STUD. 575, 581 (1988).

85 *See* Cheney-Lippold, *supra* note 72, at 167; Fourcade & Healy, *supra* note 42, at 292.

text.[86] Numerical transformations are always value-laden, are never deterministic in any objective sense, and always depend upon human judgment.[87] Even among the measurements made in "objective" or "natural" sciences, there are no neutral facts, only judgments.[88] Every scientific measurement is based upon a human decision that some things that will be included in the measurement matter, and other things that will be ignored or excluded from the measurement do not matter.[89] At the very least, some human decision about data suitability must be made. Neither do algorithmic scores provide windows on naturally occurring, immanent, or objective phenomena. They are rather the very human products of intricate, convoluted, and contrived interpretation.[90] Even slight or unnoticed changes in the criteria for measurement, or on the mix of inputs relied upon for assessment, may substantially change the outcome.[91]

The intervention of human judgment can be readily seen from the description above of predictive data processing. Machine learning does not operate by means of inductive logic, deriving general rules from specific instances, nor by means of deductive logic, deriving from application of general rules.[92] It rather operates by means of a process closest to abductive logic, that is the generation and testing of hypotheses.[93] Via iterative correlative pattern matching, machine learning systems abductively infer facts, rules, and hypotheses to speculatively explain unknown phenomena.[94] Such speculative inference requires selection and culling of the multiple relational rules generated by the system.[95] From all the myriad possible correlations detected within the data, the system preserves "best" or most "interesting" correlations to elevate to operational models.[96] This of course requires some notion of what constitutes best or most interesting, and that notion is not supplied by the machine—it has to be supplied by the machine's engineers.[97] Human judgment is required to adjust the output against both

86 David J. Cooper & Michael J. Sherer, *The Value of Corporate Accounting Reports: Arguments for a Political Economy of Accounting*, 9 ACCT. ORGS. & SOC'Y 207, 208 (1984).

87 Thus, the "facially neutral algorithm" that Professor Chander postulates in his analysis of data biases is not merely a contradiction, but an impossibility. *See* Chander, *supra* note 45, at 1036 (critiquing potentially disparate impact of "facially neutral algorithms").

88 J. BRONOWSKI, SCIENCE AND HUMAN VALUES 88–91 (Harper & Row rev. ed. 1965); *see also* PORTER, *supra* note 27, at 7 (arguing that scientific objectivity is a cultural construct).

89 *See* J. BRONOWSKI, A SENSE OF THE FUTURE: ESSAYS IN NATURAL PHILOSOPHY 6, 11 (Piero E. Ariotti ed., 1977).

90 Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1260–63 (2008).

91 Fourcade & Healy *supra* note 42, at 289.

92 Parisi, *supra* note 27, at 92.

93 *See* Peter Lipton, *Inference to the Best Explanation*, *in* A COMPANION TO THE PHILOSOPHY OF SCIENCE 184, 184 (W.H. Newton-Smith ed., 2000) (explaining hypothesis generation via abductive reasoning).

94 Parisi, *supra* note 27, at 108.

95 *Id.* at 109–10.

96 *See id.* at 110.

97 *See* Mackenzie, *supra* note 1, at 438.

overfitting and underfitting, deciding whether the fit between the model and the data is too tight so as to miss relevant correlations or too loose so as to produce spurious correlations.[98]

Machine learning is thus an iterative process requiring continual human intervention to revise, adjust, and optimize the learning process.[99] Such "artificial intelligence" is effectively a prosthetic extension of human judgment, entailing all the messy and imprecise selectivity that human judgment entails. The impact of such systematically selective or incomplete algorithmic inputs is well known from a variety of past studies. One extensively studied instance is found in the example of American credit scoring practices. This is among the best characterized and examined algorithmic profiling systems.[100] In this system, records of various personal activities, such as mortgage payments, credit inquiries, available unsecured credit lines, and length of credit history are quantified and agglomerated in a proprietary formula.[101] The best known and most widely used algorithmic credit metric is the "FICO" score produced from consumer records by the Fair Isaac Corporation.[102]

Careful examination of credit scoring practices repeatedly demonstrates that the most vulnerable and impoverished sectors of society are often those least able to avoid harmful surveillance and value extraction, but who remain invisible to advantageous classification and scoring systems.[103] It is well known that the criteria chosen for calculating the credit score render certain financial activities algorithmically visible and others invisible. Specifically, lower income individuals are systematically underrepresented or misrepresented in credit scoring because their activities are often absent from the records drawn from conventional banking, mortgage loan, credit reporting, and other financial transactions lying outside their means.[104] Other information that might be relevant to financial stability and lending risk is typi-

98 *See* TREVOR HASTIE, ROBERT TIBSHIRANI & JEROME FRIEDMAN, THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION 38 (2d ed. 2009)

99 *See* Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, COMMN'CS. ACM, Oct. 2012, at 78, 84.

100 *See, e.g.*, Marion Fourcade & Kieran Healy, *Classification Situations: Life-Chances in the Neoliberal Era*, 38 ACCT. ORGS. & SOC'Y 559 (2013); Mark Kear, *Playing the Credit Score Game: Algorithms, 'Positive' Data and the Personification of Financial Objects*, 46 ECON. & SOC'Y 346 (2017); Donncha Marron, *'Lending by Numbers': Credit Scoring and the Constitution of Risk Within American Consumer Credit*, 36 ECON. & SOC'Y 103 (2007); Martha Poon, *Scorecards as Devices for Consumer Credit: The Case of Fair, Isaac & Company Incorporated*, 55 SOCIO. REV. 284 (2007); *see also* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014) (summarizing the history and impact of consumer credit scoring).

101 Marron, *supra* note 100, at 111, 115; Poon, *supra* note 100, at 294–95.

102 *See* Poon, *supra* note 100, at 285.

103 *See* Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55, 57 (2013); Kaveh Waddell, *How Big Data Harms Poor Communities*, THE ATLANTIC (Apr. 8, 2016), https://www.theatlantic.com/technology/archive/2016/04/how-big-data-harms-poor-communities/477423/.

104 Fourcade & Healy, *supra* note 100, at 565.

cally filtered out or ignored. Financial activity common to lower income populations, such as rental payments or payday loans, is excluded from the ambit of the system.[105]

Credit scoring is representative of the reproduction of social context within algorithmic metrics. Underprivileged segments of the population generate data profiles disproportionately drawn from interactions with public assistance, law enforcement, and penal institutions. At the same time, data on more affluent populations is drawn from a different set of recording systems, yielding not only markedly different types of data records, but records that are turned to purposes that reinforce their own privilege.[106] Favored metrics are valorized, and disfavored metrics become invisible.[107] As a result of such algorithmic flattening, credit scoring, which was originally intended to help neutralize bias in lending, serves to enforce class divisions by recapitulating the circumstances of the profiled data subjects.

Further, to the extent that lower income is equated with race, differential scoring of impoverished communities effectively means that racial minorities are systematically underrepresented or misrepresented in the scoring system.[108] Some attempts to establish or "repair" credit have focused integrating marginalized "risky" populations into the FICO system by instructing them on the activities and that would be favorably recognized by the credit scoring algorithm.[109] Other attempts have focused on integrating into credit scoring the data from social systems and institutions, such as payday loans lenders, that might be characteristic of otherwise invisible populations.[110] But whether fully invisible or partially invisible, minority populations remain algorithmically differentiated from affluent populations by virtue of data perception, escaping notice because of the selective inclusion or exclusion of data within algorithmic metrics.

A distinct but related form of systemic bias stems from decontextualized reporting that does not necessarily reflect the circumstances of the individual tied to the data profile. For example, algorithmic credit scoring incorporates and heavily weights evidence of failure to make credit payments on time.[111] Even assuming the record of late payments is "accurate" rather than mistaken, in the sense of reflecting an actual occurrence, the data typically leaves little room for context or explanation. Late payments may be the result of a wide range of contextual factors, from misdirected postal mail to natural disasters. In some cases, the failure of timely payments may be due to the intrinsic characteristics or behaviors of the individual, such as forgetfulness, profligacy, or poor funds management. In other cases, the cause may be purely external and outside the individual's control, such as the loss of

105 Kear, *supra* note 100, at 360; Marron, *supra* note 100, at 125.

106 Kear, *supra* note 100, at 360.

107 Fourcade & Healy, *supra* note 100, at 565.

108 PASQUALE, *supra* note 3, at 41.

109 Kear, *supra* note 100, at 360.

110 *Cf. id.* at 361.

111 Kear, *supra* note 100, at 365 n.3.

employment, unexpected illness or injury, or identity theft. But all of these circumstances are acontextually reported and weighted similarly.

Consequently there is only an unclear or tenuous factual linkage between the living individual to whom a credit score is attached and that individual's algorithmically constructed "credit double" intended to represent the credit worthiness of the profiled data subject. Any linkage is correlative and categorical. Some instances of reported late payment may reflect systemic, underlying risks of extending the credit to a given individual; others reflect serendipitous or unique occurrences that indicate little about the individual's behavioral propensities. The ability of the individual to "correct" her credit record in the sense of adding context is extremely limited; the system makes little allowance for contextual nuance. To the extent that there is any consumer recourse to address atypical or nonrepresentative payment situations, it typically lies in convincing a lender not to report a late or missed payment, rather than any ability to add context to the report once it has entered the scoring system.

Thus, the operation of the credit scoring system offers a detailed illustration of the social impact attending the algorithmic processes now being advanced via the proliferation of machine learning systems. Identities that are incorporated within a given population become tethered to the categories as defined by the algorithm.[112] Individual identity thus increasingly depends upon algorithmic mediation of the relationship between surveillance data and categorical meaning of that data.[113] Algorithmic prediction depends on classification, and classification depends on the existence of identifiable class attributes.[114] The process of algorithmically identifying categories, such as gender, class, or preference increasingly serves to define the characteristics of gender, class, or preference.[115] Such actuarial technologies categorize individuals according to statistical distributions.[116] Categories are assigned on the assumption that profiles sharing characteristics derived from past correlations will share relevant characteristics in the future.[117]

Consequently, although algorithmic data profiling practices may be advanced as vehicles for "personalization," it should be clear that they are in fact anything but personalized. To the contrary, as described above, algorithmic processing flattens and distorts the depiction of the individual by the construction of data doubles that are reduced to portable numerical representations.[118] Those profiles are then slotted into predictive categories that are matched to selected correlative patterns. Neither the profile nor the category reflects more than the particular pattern that the algorithm deter-

112 Cheney-Lippold, *supra* note 72, at 176.

113 *Id.* at 178.

114 Mackenzie, *supra* note 1, at 433; Yeung, *supra* note 3, at 512.

115 Cheney-Lippold, *supra* note 72, at 165.

116 Jonathan Simon, *The Ideological Effects of Actuarial Practices*, 22 LAW & SOC'Y REV. 771, 772 (1988).

117 Lury & Day, *supra* note 77, at 23.

118 *See supra* notes 72–81 and accompanying text.

mines to be of interest. As Jonathan Simon puts it, such actuarial practices do not make people up, they "unmake them," offering a quantified substitute for intersubjectivity, identity, or self-awareness.[119]

Far from offering more personalized law, the data-driven metrics espoused by recent commentators may be seen as offering radically *depersonalized* law, at least in the sense of respecting the individual as a person.[120] Such practices might be said to produce more granular[121] or atomized legal metrics.[122] But particularized legal functionality is a far cry from personalization in any humanistic sense.[123] The numerical cartoons to whom such metrics are applied are artificially constructed caricatures intended for numerical manipulation rather than for individualized respect.[124]

## II. Algorithmic Reflexivity

I have to this point described how data analytics winnows and reduces the otherwise incomprehensible and unmanageable reams of data that can be gathered and stored from surveillant sources.[125] I have additionally described how the process of data intake, treatment, and categorization tethers opportunities and future behaviors to the digested vestiges of previous behaviors.[126] These information processes reflect more general social processes that are routinely engaged in quantifying and managing information. In particular, algorithmic informational processing can be tied to well-studied reflexive social processes such as commensuration.[127] A close review of such processes is foundational to understanding the social impact of algorithms generally, and the implications of algorithmic legal metrics in particular.

Commensuration, whether automated or not, occurs in the transformation of perceived *qualities* into *quantities*, creating a shared metric among the

---

119 Simon, *supra* note 116, at 792; *cf.* Hacking, *Making Up People*, *supra* note 28, at 223 (discussing the deconstruction of individual identities via statistical governance).

120 *See* Rouvroy, *supra* note 57, at 128 (observing that ostensibly personalized algorithmic profiling generates categorical sorting rather than recognizing the "unique complexity" of any human being).

121 *See. e.g.*, Christoph Busch, *Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law*, 86 U. Chi. L. Rev. 309, 314 (2019) (referring to personalized legal rules as "granularized").

122 *See id.* at 329 (referring to personalized disclosure rules as "atomiz[ed]").

123 *See* Lury & Day, *supra* note 77, at 19 (noting the gap between algorithmic personalization and individuation).

124 *See* Ian Hacking, *How Should We Do the History of Statistics?*, *in* The Foucault Effect: Studies in Governmentality 181, 193 (Graham Burchell, Colin Gordon & Peter Miller eds., 1991).

125 N. Katherine Hayles, How We Think: Digital Media and Contemporary Technogenesis 230 (2012).

126 Cheney-Lippold, *supra* note 72, at 169.

127 *See* Wendy Nelson Espeland & Mitchell L. Stevens, *Commensuration as a Social Process*, 24 Ann. Rev. Sociology 313 (1998).

objects of the transformation.[128] In doing so, this common social process simplifies information, decontextualizes information, recategorizes information, and ultimately alters social perceptions by directing attention away from certain relations and toward others.[129] This in turn alters institutional and individual behaviors toward the objects of commensuration. Rather than directly changing behavior, commensuration alters cognition, shifting attention to favored information and away from disfavored information.[130]

Commensurate perceptions are produced by means of a suite of closely entangled operations. First, commensuration simplifies information by reducing the characteristics of measured entities to quantified representations.[131] Difference in kind is represented as a difference in degree or magnitude.[132] Some information is discarded and other information is valorized in this process of quantification.[133] Complex concepts become a single number or metric that is easier to remember, easier to compare, easier to discuss and deploy in new circumstances. Commensuration thus immediately changes perceptions of information, emphasizing measured, quantifiable relationships while suppressing and eclipsing others.[134]

Second, commensuration categorizes and restructures information. Quantified metrics are grouped together and related to one another. By means of metric categorization, commensurate representations create the perception of commonality, while eliminating from consideration other differences or similarities.[135] Thus, additional information deemed extraneous or irrelevant is discarded as commensurate concepts are connected. By decontextualizing and excluding certain information so as to simplify the remaining information to a shared metric, the commensuration process makes accessing and processing information easier.[136] This also makes the final result appear more authoritative, by obscuring and removing the complicated assumptions, discretion, and arbitrariness that infuse any information output.[137]

Once broken down, this process can be readily seen in numerous well-documented examples, such as the previously described example of credit scoring. We have already seen how credit scoring assembles disparate, decontextualized measurements—payment histories, income, number of credit accounts, type of credit accounts, and so on—and reduces these to numbers that represent differences in degree among the various metrics. Some data is discarded, some ignored, some distinguished. Favored data is

128 Wendy Nelson Espeland & Michael Sauder, *Rankings and Reactivity: How Public Measures Recreate Social Worlds*, 113 AM. J. SOCIO. 1, 16 (2007).

129 *Id.*

130 *Id.*

131 Espeland & Stevens, *supra* note 127, at 316.

132 *Id.* at 316–17.

133 *Id.* at 317.

134 Espeland & Sauder, *supra* note 128, at 16.

135 *Id.*

136 *Id.* at 17.

137 *See generally* JAMES G. MARCH & HERBERT A. SIMON, ORGANIZATIONS (1958).

combined into a final commensurate score. Rather than a messy history of financial transactions, a single pristine metric appears, simplified and portable, and elevated to a position of control over the subject's opportunities or "life chances."

And, as described above, such commensurate metrics are reinforced by their own selectivity. In response to credit scoring practices, consumers who are sufficiently affluent to engage in favored transactions are shunted toward the institutions and practices that will generate records of such activity. Indeed, a variety of commercial and nonprofit institutions have grown up, offering guidance regarding the complexities of credit scoring commensuration.[138] Such organizations will assist consumers in building a favorable credit history, or will advise consumers on activities that will convert an unfavorable score to one that is more favorable.[139] In many instances, the advice or activity most amenable to appeasing the credit algorithm is not necessarily beneficial to the consumer—for example, because length of credit history is an important factor in credit scoring, consumers may be advised not to close older credit accounts, even if those accounts are unused or unneeded. Similarly, because total available credit is an important scoring metric, credit score guidance may advise against closing any available accounts even if they are no longer needed (and possibly create greater exposure to identity theft).

A second well-studied example of commensurate algorithmic practices, familiar to many readers of this Article, is the yearly practice of commensurately ranking American law schools in the publication *U.S. News and World Report* (USNWR).[140] In this process, various qualities of law schools are first quantified—reputation, bar passage rate, student quality, and other characteristics are converted to quantities that are added together in a proprietary mathematical calculus.[141] The various chosen metrics, which may have little causal or conceptual relationship to one another, are placed in a complementary and comparative relation to one another by means of a weighted formula. Other information about a school—the diversity of student body or faculty, volunteer pro bono service, student satisfaction—is ignored or discarded in the ranking procedure.[142] The resulting ranking metric, and the

138 *See* Fourcade & Healy, *supra* note 100, at 565.

139 Kear, *supra* note 100, at 350–52.

140 *See* Michael Sauder & Wendy Nelson Espeland, *The Discipline of Rankings: Tight Coupling and Organizational Change*, 74 AM. SOCIO. REV. 63 (2009). Although USNWR ranks only accredited U.S. law schools, academic institutions around the world experience similar effects from various forms of commensurate ranking. *See* Janet Chan, Fleur Johns & Lyria Bennett Moses, *Academic Metrics and Positioning Strategies*, *in* METRIC CULTURE: ONTOLOGIES OF SELF-TRACKING PRACTICES 177 (Btihaj Ajana ed., 2018).

141 Espeland & Sauder, *supra* note 128, at 10; Robert Morse, Ari Castonguay & Juan Vega-Rodriguez, *Methodology: 2021 Best Law Schools Rankings*, U.S. NEWS & WORLD REP. (Mar. 16, 2020), https://www.usnews.com/education/best-graduate-schools/articles/law-schools-methodology.

142 Wendy Espeland & Michael Sauder, *Rankings and Diversity*, 18 S. CAL. REV. L. & SOC. JUST. 587, 604 (2009).

known inputs fueling the metric, become associated with the school, shaping perceptions of the school.[143]

As in the credit scoring example, behavior is changed by the realignment and decontextualization of the quantified characteristics in law school rankings. For instance, the quality and quantity of applications to a ranked school are dramatically affected by student reliance on the score.[144] Moreover, the schools' programs, organization, and expenditures related to the defining metrics are unavoidably and profoundly altered.[145] Widespread acceptance of the ranking system creates strong incentives to invest in the measured characteristics, such as the provision of deep discounts or "merit scholarships" to students with desirable test scores.[146] The rankings generate little incentive to invest in disfavored characteristics, such as volunteer pro bono activity—indeed, such expenditures are effectively penalized for diverting resources away from the favored metrics.[147]

Even though the law school ranking and credit scoring examples rely on algorithmic processing that does not necessarily involve machine learning, the general cognitive and social characteristics of commensuration remain the same in the AI context. It should thus be clear from this summary that algorithms as described above extend, instantiate, and perpetuate the processes of commensuration that are at work in society generally.[148] Indeed, the algorithmic processes described above can be seen to act as a type of technical prosthesis for the more general social process of commensuration, extending the properties of human capacity.[149] Algorithmic scoring quantifies, simplifies, categorizes, and relates different datasets together as part of its operation, but the numerical output of the algorithmic process then serves as a starting point for further social commensuration. And, as with every other social process embodied in computer technology, the speed, scope, and effect of commensuration are magnified and enhanced by passage through the machine.[150]

Thus, algorithms incorporate and extend a larger set of reflexive social practices, defined to include any social response to a measurement that changes the object being measured.[151] An additional reflexive social effect associated with algorithmic processing is the generation of self-fulfilling

143 Sauder & Espeland, *supra* note 140, at 72.

144 Michael Sauder & Ryon Lancaster, *Do Rankings Matter? The Effect of* U.S. News & World Report *Rankings on the Admissions Process of Law Schools*, 40 LAW & SOC'Y REV. 105, 122–24 (2006).

145 Sauder & Espeland, *supra* note 140, at 71.

146 Espeland & Sauder, *supra* note 142, at 601–602.

147 Sauder & Espeland, *supra* note 140, at 73–74.

148 Espeland & Stevens, *supra* note 127, at 315–16.

149 *Cf.* Clive Lawson, *Technology and the Extension of Human Capabilities*, 40 J. THEORY SOC. BEHAV. 207 (2010) (reviewing arguments regarding technology as amplifying or supplementing bodily functions).

150 *See* LAURA J. GURAK, CYBERLITERACY: NAVIGATING THE INTERNET WITH AWARENESS 29 (2001).

151 *See* Espeland & Sauder, *supra* note 128, at 3.

prophecies. This occurs when the algorithm changes social activity and perceptions in a way that confirms its own analytical output or prediction by means of performative loops: the algorithm does not merely reinforce, but actually recreates existing social opportunities or liabilities.[152] For example, algorithmic scoring is known not merely to correspond to, but to reinforce differences in wealth, class, education, and race.[153] Scored individuals become trapped in positive or negative feedback cycles as their scores prompt behaviors that validate the score.

As illustrations, I turn again to well-documented examples, such as that of American credit scoring. Impoverished or marginalized social groups who systematically receive lower credit scores are deemed by virtue of the score to constitute a greater credit risk, resulting in more stringent terms for credit, at higher interest rates. When imposed upon a population that is already impoverished, such lending strictures naturally contribute in a higher rate of default, substantiating the algorithm's prediction of credit risk.[154] Thus credit scores do not merely predict default, but actually facilitate default.[155] The score does not simply indicate or reflect objective risk; rather, when actively employed as tool to mitigate risk, the scoring algorithm fulfills its own prediction.[156]

Similarly, so-called predictive policing analyzes inputs deemed relevant to criminal activity in order to optimize the deployment of law enforcement resources in neighborhoods that have been algorithmically calculated to offer the highest risk of future crime.[157] But deployment of such resources according to the algorithm affects the incidence of crime encountered in the areas identified by the algorithm.[158] This should not be surprising; the presence of more personnel and higher degrees of surveillance in the location naturally contributes to the detection and apprehension of misconduct, fulfilling the algorithm's prediction of heightened criminal activity.[159]

We might similarly expect, in the practice described above of predictive scoring for child protective services,[160] that the algorithmic inspection, scrutiny, and judgment of parents predicted to place their children at risk would produce stressed and irritable parents who are more likely to become angry or make mistakes. This may in fact place their children at greater risk, prompting the need for intervention of protective state institutions. As in all of these examples, the algorithmic metric is not a detached or neutral evalua-

---

152 *See* Citron & Pasquale, *supra* note 100, at 18; Fourcade & Healy, *supra* note 100, at 570.

153 *See* PASQUALE, *supra* note 3, at 15, 38; Haggerty & Ericson, *supra* note 33, at 618.

154 Fourcade & Healy, *supra* note 100, at 567.

155 *See* Akos Rona-Tas, *The Off-Label Use of Consumer Credit Ratings*, HIST. SOC. RSCH., no. 1, 2017, at 52, 62.

156 *See* Citron & Pasquale, *supra* note 100, at 15.

157 Moses & Chan, *supra* note 7, at 808; Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 113–14 (2017).

158 *See* Moses & Chan, *supra* note 7, at 810.

159 *Id.*; Selbst, *supra* note 157, at 141.

160 *See supra* note 5 and accompanying text.

tion system. The algorithmic system is instead embedded in the social systems it is deployed to predict or assess, and so will certainly perturb the function of those systems, just as those systems perturb it.

## III. Algorithmic Markets

Although the reflexive characteristics of algorithmic profiling are dramatic, they are only one example among numerous known socially reflexive practices that facilitate behavior conforming to their own predictions.[161] Other commensurate practices are known for creating social outcomes that conform to the expectations embedded in the metrics intended to measure them.[162] Perhaps the best-characterized examples of such effects are found in detailed studies of financial markets, where predictive algorithms alter not only behavior but in fact alter the social structures that produce such behaviors.[163] In particular, due to characteristics such as numerical quantification and description, economic institutions display strongly commensurate and self-fulfilling prophetic qualities. These qualities are likely to be reinforced when coupled with algorithmic profiling; thus, the most notable displays of the recursive impact of predictive algorithms have been documented in these areas.[164]

The impact of predictive algorithms in finance offers not only a graphic illustration of recursive processes in action, but an illustration that is profoundly entangled with the problem of legal metrics because the logic of financial algorithms is carried directly over to legal algorithms. First, proposals for personalized legal metrics adopt the technology and techniques for market personalization in determining legal status, and carry with them the characteristics of market personalization.[165] Indeed, many of the areas where algorithmic legal metrics have been proposed, such as contract, tort, or intellectual property law, are legal domains where private market ordering is typically part of the expected operational rationale, which animates the translation of market mechanism into legal institutions.[166] Second and not coincidentally, because of decades of pervasive economic analysis of law in the United States, both theory and doctrine of American jurisprudence are thoroughly permeated with the assumptions that animate such economic analysis, so that even public ordering is typically conceived of under paradigms of economic rationality. And third, economic rationales share with machine learning certain problematic assumptions that reinforce one another when the two intersect.

161 Espeland & Sauder, *supra* note 128, at 11.

162 *Id.*

163 *See* Mackenzie, *supra* note 1, at 442.

164 *Id.*

165 *See* Cohen, *supra* note 18, at 66; Yeung, *supra* note 18, at 264–65.

166 *See* Cohen, *supra* note 18, at 66.

### A. *Algorithmic Performativity*

As an explanatory matter, it may be most helpful to expand on the third point first: the shared assumptions underlying both economic and predictive algorithmic modeling. In many senses, the juxtaposition of algorithmic scoring with neoclassical economic models seems a natural ideological match. As I have pointed out in previous work, the nascent literature on personalized law relies upon algorithmic technological intervention as *deus ex machina*—both literally and figuratively—to instantiate the neoclassical pipe dream of zero transaction costs and perfect information.[167] Ubiquitous surveillance systems are envisioned to provide boundless data inputs; transcendent machine learning is envisioned to predict optimal regulatory responses. Such prognostications vastly exceed the engineering capacity of any feasible future implementation,[168] but are fairly typical of what has been termed a "magical worldview" that underestimates the costs and overpromises the results of AI implementation.[169]

But the algorithmic metric proposals considered here rest on an even deeper set of congruencies between economic and algorithmic practice. There is, as we shall see, considerable compatibility between the analytical abstraction entailed in algorithmic profiling and the economic abstraction of the rational actor on which personalized law proposals are premised.[170] For example, algorithmic profiling seems "scientific" in the sense that it is expressed as numerical output that purports to objectively measure underlying patterns of social activity, in much the way that investigation of the physical world purports to objectively measure underlying patterns of material activity.[171] Economics, too, has long relied on analogous (and dubious) assumptions of objective observation in developing its analytical apparatus, particularly those offshoots of economic analysis that have taken root within the legal academy.[172]

Similarly, the collection and integration of data profiles for predictive metrics relies upon the idea that they reflect an underlying rational, efficient, and knowable actor whose preferences can be forecast and manipulated in a defined and predictable fashion.[173] This postulated actor would be perfectly

167 *See* Burk, *supra* note 26, at 288–89.

168 *See* Cooper & Levy, *supra* note 20.

169 *See* Elish & boyd, *supra* note 41, at 73–74; *see also* Malcolm Campbell-Verduyn, Marcel Goguen & Tony Porter, *Big Data and Algorithmic Governance: The Case of Financial Practices*, 22 NEW POL. ECON. 219, 220 (2017) (labeling as "techno-utopian" the optimistic view that algorithmic governance will "overcome the imperfections of politics and faulty forms of knowledge"); Mackenzie, *supra* note 1, at 436 (describing the romantic notion that machine learning will transcend unruly and chaotic data inputs).

170 *See* Fourcade & Healy, *supra* note 41, at 20.

171 *See* Mackenzie, *supra* note 1, at 433 (noting that machine learning processes assume the existence of stable and distinct external categories).

172 *See* Jeanne L. Schroeder, *Rationality in Law and Economics Scholarship*, 79 OR. L. REV. 147, 149–50 (2000).

173 Fourcade & Healy, *supra* note 41, at 20.

at home in neoclassical economic modelling. The proposal that consumer profiling can be established via predictive data analysis assumes that there exists a static and fixed set of consumer preferences that can be derived from the data trails that individuals leave in their wake, much as within the structure of economic modelling, the neoclassical premise of the "rational actor" constitutes a parallel assumption regarding a defined subject with predictable, intrinsic categorical preferences. Deployment of algorithmic prediction assumes that there are intrinsic preferences present to be detected, as does the assignment of consumers to a position on the economists' demand curve.[174]

To some degree these synergies between economic rationales and algorithmic design may, when conjoined, amplify the most dubious features of each. The premise of algorithmic divination of intrinsic consumer preferences should already be at least suspect, if not entirely discredited; we have seen in the discussion to this point that algorithms cannot be said to discover or extract intrinsic quantitative properties. To the contrary, they reflexively shape the outcomes that they predict or report.[175] On the economic side, I will not belabor here the dubious premise of the rational actor—despite its surprising persistence, this has been shown repeatedly to constitute a serious deficiency in policy analysis and prescriptions drawn from economics.[176] But as Yochai Benkler notes, the concept of exogenous, freestanding preferences has never been a reasonable assumption in contemplating consumer behavior, and this fiction reaches new levels of absurdity in the face of algorithmically profiled markets.[177] Indeed, the absurdity is enormously amplified when the electronic speed and reach of algorithmic systems is engaged.

However, having already outlined above the reflexivity effects attending algorithmic profiling, I need now to underscore the parallel reflexivity problems attending neoclassical economic premises that to a significant degree animate the proliferation of automated algorithmic metrics. As Stephanie Bair has pointed out in some detail, rationality and the rational satisfaction of preferences show problematic signs of behavioral reflexivity.[178] What we deem rationality is neither fixed nor immutable, but rather changes with the circumstances in which an individual finds him or herself.[179] Preferences are not static, but dynamic. It is not simply that the individual may gain access to better or worse information on which personal preferences depend, nor that the individual's attitude may change depending on the situ-

174 *Id.* at 23.

175 *See supra* notes 148–53 and accompanying text.

176 *See, e.g.*, Geoffrey M. Hodgson, *On the Limits of Rational Choice Theory*, ECON. THOUGHT, no. 1, 2012, at 94, 94; Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CALIF. L. REV. 1051 (2000); Thomas S. Ulen, *Rational Choice Theory in Law and Economics, in* 1 ENCYCLOPEDIA OF LAW AND ECONOMICS 790 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000).

177 Yochai Benkler, *Degrees of Freedom, Dimensions of Power*, DAEDALUS, Winter 2016, at 18, 23–24.

178 Stephanie Plamondon Bair, *Malleable Rationality*, 79 OHIO ST. L.J. 17, 19–20 (2018).

179 *Id.* at 20.

ation: the brain actually rewires its neural pathways depending upon environmental stresses and experiences.[180] "Rationality" is thus the product of social circumstances that we create; law and policy shape our attitudes and preferences at a biological level. When we base social policy on "rational" preferences we may be dynamically altering or cementing the preferences on which the policy supposedly depends in a mutable feedback mechanism.[181]

On a broader scale, such recursive effects are even more troublesome with regard to the concept of the economic "market." Because of its distinctive character, economics may display some of the strongest reflexive social effects entailing both commensuration and self-fulfilling prophecies.[182] Economics routinely reduces social relations to quantified numerical and acontextual representations. Unlike physical phenomena, which to be studied must be instrumentally quantified, the subject matter of economic analysis is *inherently* quantified, in the form of numerical data drawn directly from the phenomena studied—corporate profit, net income, or market surplus entail numerical representations that are intrinsic to the activities of corporations, income, or markets.[183] These commensurate concepts are highly portable because of their numerical representation; because they have been rendered context-free and appear authoritative, such quantified representations can be quickly redeployed into new contexts and may quickly become adopted outside their initial circumstance.[184]

These numerical representations are in turn the quantities that are affected and manipulated by the economic models that process them. As a consequence, there is evidence that, far from describing preexistent or immanent phenomena, economic models serve to induce and engender the phenomena that they purport to describe.[185] This *performativity* thesis, as articulated by Callon and advanced by subsequent researchers, postulates that the formal models and techniques developed in economics are responsible for the market characteristics they compose.[186] Economic models, like

180 *See* Richard J. Davidson, Katherine M. Putnam & Christine L. Larson, *Dysfunction in the Neural Circuitry of Emotion Regulation—A Possible Prelude to Violence*, 289 SCI. 591, 594 (2000). *See generally* Todd A. Hare, Colin F. Camerer & Antonio Rangel, *Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System*, 324 SCI. 646 (2009); Todd F. Heatherton & Dylan D. Wagner, *Cognitive Neuroscience of Self-Regulation Failure*, 15 TRENDS COGNITIVE SCIS. 132 (2011).

181 *See* Bair, *supra* note 178, at 33–34.

182 *Cf.* Espeland & Sauder, *supra* note 128, at 11–12, 16.

183 Duncan K. Foley, *The Strange History of the Economic Agent*, 1 NEW SCH. ECON. REV. 82, 82 (2004).

184 Kear, *supra* note 100, at 347.

185 Donald MacKenzie, Fabian Muniesa & Lucia Siu, *Introduction*, *in* DO ECONOMISTS MAKE MARKETS? ON THE PERFORMATIVITY OF ECONOMICS 1, 4–5 (Donald MacKenzie, Fabian Muniesa & Lucia Siu eds., 2007); Kieran Healy, *The Performativity of Networks*, 56 EUR. J. SOCIOLOGY 175, 177–78 (2015).

186 Michel Callon, *Introduction: The Embeddedness of Economic Markets in Economics*, *in* THE LAWS OF THE MARKETS 1, 2 (Michel Callon ed., 1998); DONALD MACKENZIE, AN ENGINE, NOT A CAMERA: HOW FINANCIAL MODELS SHAPE MARKETS 12–20 (2006); Donald MacKenzie &

other human technologies—whether wheels, or lenses, or amplifiers, or transmitters, or antibiotics that serve as prosthetics to extend and enhance endowed human capabilities—appear to serve as "cognitive prosthetics" allowing their users to engage in calculative techniques demanded *by* the models.[187] Thus, in at least some cases, and possibly with some frequency, economic models in fact *perform* the theories they articulate, creating rather than discovering the characteristics of the market relationships that are the focus of disciplinary study.

Such performativity has been tied to a range of effects in the marketplace, from merely conceptual changes to complete restructuring of markets.[188] At the most modest level, dubbed generic performativity, economic theories or concepts are adopted by market actors other than economists, changing those actors' views on market policy and activity. Economic concepts that have a more substantial impact on the market by altering practices may be labeled as effective performativity. The very strongest performativity effects, known as Barnesian performativity,[189] creates a type of self-fulfilling prophecy or substantive feedback loop in which the application of the theory produces its own predictions.[190]

In this strongest, Barnesian form, performativity effects are not merely behavioral changes arising from the ideological indoctrination into the assumptions embedded in the models, although certainly this may occur.[191] For example, behavioral economics has shown that individuals who are exposed to undergraduate economics theories become more selfish, or "rational," in line with the expectations of those theories to which they are exposed.[192] Certainly the implementation of predictive economic models changes the behavior of market participants in such ways.[193] But Barnesian performativity effects are not merely generic or effective changes in attitude or behavior; rather, they alter social market practices on a wide scale. Thus, when transported into financial practice, the capabilities arising from economic models can cause markets to increasingly resemble the assumptions, behaviors, and predictions embedded in the models.[194]

---

Yuval Millo, *Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange*, 109 AM. J. SOCIO. 107, 107–108 (2003).

187 Michel Callon & Fabian Muniesa, *Economic Markets as Calculative Collective Devices*, 26 ORG. STUD. 1229, 1238 (2005). *See* Lawson, *supra* note 149 (discussing technology as extending or supplementing bodily capabilities).

188 MACKENZIE, *supra* note 186, at 18–19.

189 *Id.* at 165–66.

190 Barry Barnes, *Social Life as Bootstrapped Induction*, 17 SOCIO. 524, 536–38 (1983).

191 *See* MACKENZIE, *supra* note 186, at 19.

192 Healy, *supra* note 185, at 194.

193 *See* Mackenzie, *supra* note 1, at 442.

194 *See* MacKenzie et al., *supra* note 185, at 6–7.

### *B. Calculated Consumers*

What I have described above regarding the performativity of markets, and the numerical models that drive such performativity, bears directly on our understanding of the effects of algorithmic legal metrics. The numerical models on which each are based are not merely value-laden, but value-implementing. The introduction of numerical models alters and shapes the social contexts in which they operate. Such effects constitute a clear contradiction to the assumptions under which numerical models are adopted in economic conception of the market, and under which such models are carried into legal or regulatory practices.

Like the general corpus of neoclassical economic work on which American law and economics is founded, proposals for algorithmic determinations are based upon the supposition that markets are neutral fora in which participants exchange whatever capital endowments they happen to possess: property, skills, or other assets.[195] Within this idealized, immanent, and spontaneous forum, participants are assumed to exchange goods according to their innate or inherent preferences. The corollary to this axiom is the assumption that, to the extent market exchanges may be biased, they reflect biases carried over from historical legacies, social attitudes, or inequalities derived from nonmarket structures.[196] To the extent that neoclassical models recognize that markets entail social relations, it tends to assume away such relations, either ignoring them entirely, or, on the rare occasions when they are acknowledged, treating them as a "transaction cost" or impediment to the otherwise ideal and frictionless operation.[197]

But on more careful consideration, it is to the contrary clear that what is deemed "the market" is actually a complex social system comprised of a conglomeration of equally complex social mechanisms for commodifying, valuing, and characterizing goods so that they can be associated and reassociated with different participants. Markets, whether digital or physical, are reliant for their function on elaborate systems of calculative practices. This is true whether the market in question is a grocery store, an open-air bazaar, an electronic trading exchange, or an internet platform. There is nothing neutral, immanent, or impassive about such practices.

For example, the characteristics of goods, which allow the good to be traded on the market, are neither inherently present nor externally

---

195 *See* 2 MAX WEBER, ECONOMY AND SOCIETY: AN OUTLINE OF INTERPRETIVE SOCIOLOGY 927–28 (Guenther Roth & Claus Wittich eds., 1978); *cf.* Richard A. Posner, *An Economic Analysis of Sex Discrimination Laws*, 56 U. CHI. L. REV. 1311, 1315 (1989) (explaining gender wage gaps on the argument that "[t]he average woman will therefore [due to family and childbearing] invest less in her human capital, causing her wage to be lower than the average man's, since a part of every wage is repayment of the worker's investment in human capital").

196 Fourcade & Healy, *supra* note 100, at 560.

197 Mark Granovetter, *Economic Action and Social Structure: The Problem of Embeddedness*, 91 AM. J. SOCIO. 481, 484 (1985).

attached.[198] They are rather constructed by the interaction of market entities via processes that render the good subject to calculation. "Calculation" in this sense has little to do with mathematics, but rather to do with the process of qualifying social relationships so as to render a given entity the object of market exchange.[199] Calculability is a fundamental characteristic of products and services in the marketplace, but since it is not an intrinsic characteristic it is must be created by the action of interconnected forces that make up the market. Objects are transformed into marketable goods by a social process of definition that imbues them with the perceived characteristics necessary for transfer from the seller's context to the buyer's context.[200]

Thus, as Callon and Muniesa have shown in a series of pathbreaking studies, markets accomplish calculation by elaborate processes of isolation, classification, and association— removing goods from one context, grouping them with other objects or persons, and associating them with a different context in order to achieve a transfer from one social situation to another.[201] Calculability includes stabilization of the goods' perceived qualities and relationships, rendering them definite and fixed so that property rights can be applied and transferred to the discrete object.[202] Objects must be rendered "calculable" in this sense in order to be transferred through the social mechanisms that comprise the market. Market processes entail a complex network of calculative agencies—labels, shopping baskets, salespeople, displays, packaging, advertisements, receipts, bar codes, wallets, debit cards, warranties, cash registers, and far more—that are engaged in rendering objects as calculable goods.[203]

Commensurate processes are fundamental to enabling such calculation. In particular, the kind of predictive analytical scores we have been considering here facilitate the determination and calculation of risky market entities. The score provides as a common reference point, a number offering the illusion of stable objective verity for an unknown quantity—future results.[204] The flattening of heterologous data creates the expectation that everything that can be reduced to bits is calculable, and the illusion that future uncertainty has been transformed to predictability.[205] This permits stable and determinative negotiation among market participants, allowing markets to function as if the future were in fact knowable.[206] Allocation of risk is then based on the existence of an entity that purports to predict risks, allowing the

---

198 Callon & Muniesa, *supra* note 187, at 1234.

199 *Id.* at 1231; *see also* Paolo Totaro & Domenico Ninno, *The Concept of Algorithm as an Interpretive Key of Modern Rationality*, THEORY CULTURE & SOC'Y, July 2014, at 29, 30 ("[T]he concept of calculation is very general and does not necessarily imply the manipulation of numerical symbols.").

200 Callon & Muniesa, *supra* note 187, at 1233–34.

201 *Id.* at 1231–32.

202 *Id.* at 1233.

203 *Id.* at 1238.

204 Kear, *supra* note 100, at 352.

205 Amoore & Piotukh, *supra* note 54, at 361.

206 Kear, *supra* note 100, at 352.

social relation of risk to become attached to a technical object—the algorithm.[207] In this fashion the algorithmic system becomes a technical prosthesis performing a calculative task—risk prognostication—that could otherwise not occur.

Thus, it has become increasingly clear that algorithmic data profiling has an intimate relationship with the partitioning, classifying, and associative social functions of markets.[208] Computer algorithms characterize, isolate, classify, aggregate, and enumerate representations of goods, including representations of human individuals, which are then framed and reframed in social spaces in accordance with the rules established by such algorithms.[209] Surveillant algorithms thus constitute calculative devices in this sense; they organize and govern processes by which market entities are rendered calculable, integrating the actions of a discrete set of calculative agencies.[210] And this effect is not separate or discrete from surveillant algorithmic processes in other contexts: the proposals to algorithmically quantify legal status propose the same systems, to the same end, which is to produce a calculable citizenry.

Understanding the functioning of algorithms in the market, and the prospect for parallel functions in the legal system, necessarily brings us to examination of the assumptions underlying algorithmic markets, which will be embedded in proposed algorithmic legal metrics. We have seen in the discussion above that although neoclassical economics assumes that markets are neutral and immanent, and that market actors are rational with endogenous preferences, in fact, both markets and the preferences of their participants are structured by social forces. Increasingly, such forces include the effects we have identified from algorithmic profiling. Profiling flattens and decontextualizes the information it processes, creating rather than simply measuring the metrics it reports. The resulting reflexive characteristics of algorithmic profiling create an environment of calculated self-surveillance that fills in the vacuum left by ostensibly neutral economic assumptions.

It is thus possible to draw a straight line from the accumulation of surveillant data through the reflexive effects of algorithmic scoring to the self-discipline of the emerging digital market structure. Given that the production of algorithmic scores reinforces self-surveillance, we may ask what this means for market behavior.[211] A serious concern is whether businesses may intentionally exploit the reflexivity effects described here in order to strategically restructure consumer preferences so as to shunt behavior into the businesses' preferred consumption pattern.[212] There is some evidence that this

207 *See id.* at 353.

208 *See* Julie E. Cohen, *The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy*, 31 PHIL. & TECH. 213, 214 (2018); Fourcade & Healy, *supra* note 41, at 10.

209 Callon & Muniesa, *supra* note 187, at 1242.

210 *See id.*

211 Fourcade & Healy, *supra* note 100, at 568.

212 *See* Gerhard Wagner & Horst Eidenmüller, *Down by Algorithms? Siphoning Rents, Exploiting Biases, and Shaping Preferences: Regulating the Dark Side of Personalized Transactions*, 86 U. CHI. L. REV. 581, 597–603 (2019); Yeung, *supra* note 18, at 263; Karen Yeung,

is occurring.[213] Of course, the algorithmic effects I have described here are expected to occur whether or not they are intended, merely by the deployment of algorithmic ranking systems, whether for wise or nefarious or entirely unanticipated purposes. But it is increasingly clear that these effects are increasingly intentional, anticipated, and largely nefarious.[214]

Thus, the current data-driven "platform economy" thrives within the surveillant assemblage, extracting value from surplus or collateral information generated by routine activities.[215] The aggregation, processing, and extraction of collateral data is enabled through situation of consumer activity within the surveillant assemblage.[216] Certain other business models, such as the fitness tracking Fitbit data device, rely even more explicitly on such self-surveillance incentives to sell devices and services.[217] These businesses deliberately leverage the reflexive nature of algorithmic quantification to induce voluntary, deliberate self-surveillance as a proxy for self-improvement, generating ongoing tracking metrics for health, learning, or other user activity. The devices or services offer commensurate scoring for the user's accomplishments, in effect creating a dynamic "accomplishment double" as a stand-in for the actual user activity.

On either such business model, the goal is to condition users to data tracking, commensuration, and reflexive response, to produce calculated consumers from which value can be extracted in the marketplace—value both in the form of payment for the self-surveillance service, and in the form of the data that can be repurposed for the benefit of the service supplier.[218] The induced pursuit of algorithmic metrics thus becomes a site for self-investment along the parameters dictated by the algorithm.[219] Classified individu-

---

*'Hypernudge': Big Data as a Mode of Regulation by Design*, 20 INFO. COMMC'N & SOC'Y 118, 119 (2017).

213 *See* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1003–18 (2014); Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 4–7 (2019); *see also* Abbey Stemler, Joshua E. Perry & Todd Haugh, *The Code of the Platform*, 54 GA. L. REV. 605, 633–38 (2020) (describing behavioral manipulation by platform businesses). Much of the documented user manipulation in platform business models involves behavioral "nudges" designed into platform architecture. *See, e.g.*, *id.* The algorithmic effects I consider here are to the contrary constitutive and pervasive.

214 *See* Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75 (2015).

215 *See* Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 140–43 (2017); Fourcade & Healy, *supra* note 41, at 11.

216 Haggerty & Ericson, *supra* note 33, at 615–16.

217 *See* Katelyn Esmonde & Shannon Jette, *Assembling the 'Fitbit Subject': A Foucauldian-Sociomaterialist Examination of Social Class, Gender and Self-Surveillance on Fitbit Community Message Boards*, 24 HEALTH 299 (2020).

218 Aristea Fotopoulou & Kate O'Riordan, *Training to Self-Care: Fitness Tracking, Biopedagogy and the Healthy Consumer*, 26 HEALTH SOCIO. REV. 54 (2017).

219 *See* Fourcade & Healy, *supra* note 41, at 20.

als are trained to become entrepreneurs of themselves[220] by absorption into the surveillant network.

The algorithmic outputs around which the market is constructed are both market derived and market oriented.[221] They create classifications, which in turn are the basis for new classes, new social divisions with life chances tied to each particular classification.[222] As an integral part of their operations, markets generate boundary classifications, where a particular individual or group of individuals is included or excluded from certain types of transactions.[223] Such processes appear to be endemic to the market, which competitively measures and sorts participants, separating and recombining them into market categories to which different economic rewards or punishments are attached, and from which profits can be most efficiently extracted.[224] Rather than reflecting or echoing social class, such processes may actually serve to create social classes.[225]

Although market mechanisms have long imposed such classification on participants, algorithmic self-surveillance accelerates and leverages such processes. Sociological analysis of market functions provides compelling evidence that markets increasingly are intentionally structured so as to stratify and create social differences by means of what have been called *classification situations*, that is, by the imposition of strategic class differentiation in a manner that alters the opportunities of those on whom the classifications are imposed, including affecting their access to markets.[226] Comparative scoring produces graduated or continuous classifications in which individuals are placed on a scale with metrics tied to ranked transaction levels: higher or lower insurance premiums, favorable or unfavorable interest rates, better or worse levels of privilege or service. These two types of classifications are never entirely separate; graduated classifications have cutoff or threshold points that demarcate categories of transactions, including complete exclusion from transactions.[227] Such profiles affect the life chances of the individuals whom they categorize, resulting in loans approved or refused, employment offered or denied, differential terms of service allowed or disallowed, access permitted or restricted, prices presented at favorable or unfavorable rates.[228]

---

220 *See* MICHEL FOUCAULT, THE BIRTH OF BIOPOLITICS: LECTURES AT THE COLLÈGE DE FRANCE, 1978–79, at 226 (Michel Senellart ed., Graham Burchell trans., 2008); *see also* Dorthe Brogård Kristensen & Minna Ruckenstein, *Co-Evolving with Self-Tracking Technologies*, 20 NEW MEDIA & SOC'Y 3624, 3625–26 (2018) (examining the characteristics of the "Quantified Self" improvement community).

221 Fourcade & Healy, *supra* note 41, at 22.

222 *Id.*

223 *See* Fourcade & Healy, *supra* note 100, at 564.

224 *Id.* at 560–61.

225 *Id.* at 561.

226 *Id.* at 560.

227 *Id.* at 564–65.

228 *See* Fourcade & Healy, *supra* note 41, at 22.

The power of such market stratification has been enormously amplified by the availability of massively pervasive data gathering and analysis systems—indeed, by precisely the kind of systems that would form the basis for any of the proposed schemes for predicting or personalizing legal status, and the implications for algorithmic personalized law are manifest in the current digital marketplace. The purpose of private-sector surveillance and algorithmic profiling is to develop systems of categories that will maximize profit extraction from the individuals profiled.[229] These categories are constructed according to the logic of extraction, and the opportunities made available to those slotted into the categories will be configured on the same basis.[230] Consumer data, which is to say consumer data doubles, and ultimately consumer activity, is rendered calculable via such classification.[231] Rather than surveilling consumers to guess at their preferences from their point of view, predictive analytics now surveils them in order to classify them according to the producer's point of view.[232]

Thus, much of the use of surveillant algorithmic scoring in the marketplace is oriented toward the creation of self-monitoring calculative selves.[233] Rather than the state discipline described by Foucault, this is a market discipline.[234] Rather than being oriented toward the imposition of social ordering, the internalized redefinition imposed by private predictive market surveillance is oriented toward exploitation and extraction.[235] Some previous commentators have recognized that the data gathered by private surveillance may differ in quality and purpose from that gathered by the state.[236] But within the current surveillant assemblage, there is little division between data profiling by the market or by the state; public and private sources are combined for algorithmic processing. And the disciplinary effect internalized by the subjects of algorithmic commensuration is common to public or private profiling.

## IV. Bias and Reflexivity

This brings us back to the problem of bias, or more properly, to the problem of inequity. The impact of calculative algorithmic self-surveillance has fallen and will continue to fall unevenly across different population segments. We have already seen that algorithmic processing cannot produce neutral profiles or objective scores, but is unavoidably biased, often in unpredictable ways. A number of thoughtful previous commentators have put sub-

229 *See id.*

230 *Id.* at 14.

231 Cohen, *supra* note 208, at 228.

232 Fourcade & Healy, *supra* note 41, at 23.

233 Fourcade & Healy, *supra* note 100, at 564.

234 *Cf. supra* note 39 and accompanying text.

235 Julie E. Cohen, *What Privacy Is For*, 126 Harv. L. Rev. 1904, 1916 (2013); Fourcade & Healy, *supra* note 41, at 25.

236 Niva Elkin-Koren & Michal S. Gal, *The Chilling Effect of Governance-by-Data on Data Markets*, 86 U. Chi. L. Rev. 403, 407–08 (2019).

stantial consideration into certain aspects of such algorithmic bias.[237] These critiques have employed the term "bias" in a variety of ways, addressing a variety of potentially discriminatory influences in algorithmic processing.[238] Oscar Gandy catalogs several such sources of bias in automated systems, ranging from imperfect data and erroneous data processing to the disparate impact of rational data processing outputs.[239]

One recurring, prominent, key concern in the existing literature on algorithmic systems is whether machine learning may reproduce social biases due to tainted or incomplete data inputs; as the old data processing maxim says, garbage in, garbage out.[240] Such bias could occur in the data used to train the system, or in the data actually analyzed through the system. Of course, data are always biased in some fashion; as Bowker famously observed, data are never raw, but are always collected and adapted to some purpose.[241] It is possible that the training or processed data might be intentionally "cooked" so as to produce socially biased results. But probably such intentional data discrimination would be rare. Rather, the primary common concern found in existing critiques of algorithmic data processing revolves around the inadvertent production of "inaccurate" profiles of its subjects.

Thus, on this view of bias, the question is whether the predictive system is "fit for purpose," that is, whether the outputs are accurate with regard to the intent or "fit" with the phenomena that the algorithm is attempting to predict. I have already discussed in this Article at least three types of algorithmic processing problems that might contribute to the accuracy concern. First, as I have shown above, data collection is never uniform or comprehensive. The sprawling expanse of the surveillant assemblage is uneven and selective, differentially harvesting information according to the caprice of class, race, geography, profession, age, and activity. Entire populations may be visible, invisible, or distorted by data perception.

Second, algorithmic scoring might be unintentionally skewed by the necessary "cooking" that Bowker identifies. The process of joining, sorting, categorizing, and analyzing ingested data as described above is unavoidably biased. By necessity the processing algorithm discards or subordinates information deemed unimportant. Critically, because machine learning systems develop and execute their own analytical criteria, whether supervised or unsupervised, it may be impossible to know exactly what information is dis-

---

237 Gernot Rieder & Judith Simon, *Big Data: A New Empiricism and Its Epistemic and Socio-Political Consequences*, *in* BERECHENBARKEIT DER WELT? PHILOSOPHIE UND WISSENSCHAFT IM ZEITALTER VON BIG DATA 85, 92 (Wolfgang Pietsch, Jörg Wernecke & Maximillian Ott eds., 2017) (summarizing critiques of data analytics).

238 *See* David Danks & Alex John London, *Algorithmic Bias in Autonomous Systems*, 26 INT'L JOINT CONF. ON A.I. 4691 (2017).

239 Oscar H. Gandy Jr., *Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems*, 12 ETHICS & INFO. TECH. 29 (2010).

240 Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 683–84 (2016).

241 *See* GEOFFREY C. BOWKER, MEMORY PRACTICES IN THE SCIENCES 184 (2005) ("Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care.").

carded or decontextualized, or exactly how the resulting output is biased as a result.[242]

These forms of bias are simply unavoidable, and the second may be humanly incomprehensible. This inevitability is to a greater or lesser extent recognized by previous commentators, who anticipate some inherent data discrimination, but who then assume that the bias problem could be fixed if the system is fit for purpose, if it accurately predicts the intended target behaviors.[243] One palliative measure frequently suggested is a kind of due process standard that would allow subjects to contest and correct their profiles so as to make the data fairer or more accurate.[244] A related suggestion is to implement a kind of "algorithmic affirmative action" that would proactively ameliorate whatever socially detrimental effects occur due to biased algorithmic outputs.[245]

Such suggestions, as indicated at the outset of this Article, are intended to address problems of governance by algorithms.[246] But they do not engage the third source of algorithmic bias identified here: the reflexive and performative nature of algorithmic metrics. In the process of machine learning, information is selected, ranked, matched, reconfigured, and evaluated according to its *social use*, and predictive algorithms then recursively look for social use correlations among decontextualized inputs.[247] The algorithmically derived rules generated by the machine learning process are ultimately determined by the social practices that generate the processed data.[248] Thus, obtaining unbiased algorithmic predictions is not a question of "cleaning up the data" or somehow filtering either training or analytical data for social biases. This insight is key: *the very notion of predicting social behaviors from correlations to past social behaviors necessarily captures whatever biases are incorporated into past social behaviors, reinscribing them into predictions about future behaviors.*

Thus, the inescapable bias in predictive algorithms stems not from inaccuracies in the data or in the data processing—although they are surely there—but from the distortions inherent in the *entire project* of predictive social correlations. Certainly, the accuracy of data profiling may be a serious and legitimate concern that must be taken into account in assessing proposals such as reliance on algorithmically generated legal metrics. But this is for the most part not my concern in this Article. To the contrary, I fear that the vilification of data profiling due to potentially mistaken or inaccurate profiles

242 *See* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.–June 2016, at 6–7.

243 *See, e.g.*, Citron & Pasquale, *supra* note 100, at 24; Zalnieriute et al., *supra* note 1, at 453–54.

244 Citron, *supra* note 90, at 1301–13; Citron & Pasquale, *supra* note 100, at 19; Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 109–28 (2014).

245 *See* Chander, *supra* note 45, at 1039–41.

246 *See supra* notes 25–30 and accompanying text.

247 *See* Parisi, *supra* note 27, at 102.

248 *See id.* at 109.

largely misses the point: algorithmic determinations are socially mediated, embedded in and determined by the social meanings attributed to their underlying processes.[249]

Consequently, I am primarily concerned here with a distinct, intractable, and inevitable set of costs in the operation of algorithmic data processing, which is that it engages a set of social mechanisms leading to reflexive quantification, manipulative self-discipline, and ultimately to calculated publics. These effects are fed by the accuracy problems identified by previous commentators, but concerns regarding reflexive effects are not necessarily concerns regarding accuracy. To the contrary, *all profiling systems, even if considered "accurate" for the purpose of their design, will produce the type of detrimental internalization feedback effects that I describe here.* Even if data profiles can be "fixed" or improved for accuracy, this does not address the set of issues I ultimately engage with in this Article.

Consider as a specific example the development of "personalized" algorithmic negligence scoring as described above.[250] Some commentators have argued that data profiling might be used to tailor tort incentives to the individual, using algorithmic scoring to adjust the standard of care required for negligence liability.[251] They argue, for example, that some data profiling research suggests an ability to algorithmically predict risky behavior, such that profiles might be considered in setting an individualized standard of care.[252] The assumption in the "personalized tort" analysis seems to be that such risk assessments measure static and inherent behavioral qualities of the individual.

Yet such a project occurs within an inescapable social framework. Recall the discussion above regarding the malleable nature of rationality and preference.[253] There is a growing body of empirical evidence supporting the conclusion that poverty changes cognitive functions, leading to suboptimal or antisocial behaviors.[254] Impoverishment narrows the cognitive focus of its victims, inhibiting long-term planning and prompting objectively imprudent activity.[255] Impoverished people perceive options and circumstances radically differently than would a wealthier person facing similar choices.[256] Substantial evidence suggests that these effects are biologically rooted, that the

249 *Id.* at 92.

250 *See supra* note 15 and accompanying text.

251 *See* Ben-Shahar & Porat, *supra* note 13, at 679–80.

252 *Id.* at 682–83.

253 *See supra* notes 178–84 and accompanying text.

254 *See* SENDHIL MULLAINATHAN & ELDAR SHAFIR, SCARCITY: WHY HAVING TOO LITTLE MEANS SO MUCH 41–42 (2013); Anandi Mani, Sendhil Mullainathan, Eldar Shafir & Jiaying Zhao, *Poverty Impedes Cognitive Function*, 341 SCI. 976, 976 (2013); Anuj K. Shah, Sendhil Mullainathan & Eldar Shafir, *Some Consequences of Having Too Little*, 338 SCI. 682, 682 (2012).

255 Shah et al., *supra* note 254, at 684.

256 *See* Anuj K. Shah, Jiaying Zhao, Sendhil Mullainathan & Eldar Shafir, *Money in the Mental Lives of the Poor*, 36 SOC. COGNITION 4, 5 (2018).

brain chemistry and neural pathways of those subjected to severe resource limitations change in response to such stresses.[257]

Thus we would expect that the classic "reasonably prudent person standard" from negligence law will be experienced differently across different social strata. What is reasonable to a person with abundant resources will be different than what is reasonable to a person with impoverished or constrained resources. The individual who has experienced severe resource scarcity may be more likely to behave carelessly, to make mistakes, to ignore or overlook precautionary measures. Taking such differences into account might be advanced as a goal of "personalized" negligence law. However, explicitly embedding personalized scoring in the imposition of tort liability can be expected to reinforce and recreate such social differences. Just as credit scoring produces "personalized" interest rates and loan terms, tailored to the algorithmic identity of the consumer,[258] so "personalized" tort remedies may be expected in a similar fashion to be distributed differentially across social classes, with different impacts on impoverished or subordinated groups.

Further, as seen in the case of existing calculative metrics such as credit scoring, we can anticipate that algorithmic legal scoring will inevitably obscure or discard certain types of information in favor of other types. Given the stark differences in data profiling across differently situated social classes, and in particular the differences in profiling according to wealth, it would undoubtedly be the case that any algorithmically generated legal metric score will entail gaps in its dataset. The poorest are the most likely to fall into such datafication blind spots, meaning for example that those least able to pay for the imposition of algorithmically determined negligence liability are those most likely to be flagged—accurately or inaccurately—by an algorithmic negligence profile.

Thus, even if an individual is "correctly" flagged as potentially negligent, reflexivity creates a strong likelihood that a data-matching score for negligence would tend to cement someone labeled with such a score into the category of "imprudent" persons. Being treated as a negligent or risky subject will reinforce and exacerbate the social stresses that led to the classification in the first place. The impoverished subject who matches the negligence profile is likely to remain impoverished due to increased financial stress stemming from the algorithmic determination itself. We might for example

257 *See supra* note 180 and accompanying text. Perhaps not surprisingly, children who grow up in impoverished circumstances show altered brain development. *See* Martha J. Farah et al., *Childhood Poverty: Specific Associations with Neurocognitive Development*, 1110 BRAIN RSCH. 166, 168 (2006); Daniel A. Hackman & Martha J. Farah, *Socioeconomic Status and the Developing Brain*, 13 TRENDS COGNITIVE SCIS. 65, 65 (2009); Kimberly G. Noble et al., *Family Income, Parental Education and Brain Structure in Children and Adolescents*, 18 NATURE NEUROSCIENCE 773, 775 (2015); Kimberly G. Noble et al., *Socioeconomic Disparities in Neurocognitive Development in the First Two Years of Life*, 57 DEVELOPMENTAL PSYCHOBIOLOGY 535, 545 (2015).

258 *See* DONNCHA MARRON, CONSUMER CREDIT IN THE UNITED STATES: A SOCIOLOGICAL PERSPECTIVE FROM THE 19TH CENTURY TO THE PRESENT 151 (2009).

expect that the individual with an unfavorable algorithmic negligence score could quickly find herself subject to higher insurance premiums (or perhaps stripped of insurance), or subject to higher civil or criminal liability for posing a social risk. While this would in some sense validate the prediction of the algorithm, the algorithm becomes a contributing factor to its own prediction.

And because commensurate rankings are highly portable, they tend to quickly migrate to contexts outside of their original setting. The unintended proliferation of algorithmic scores is apparent, for example, in the increasing incidence of "off-label" uses of credit scores outside of credit decisions, extending to employment vetting, security screening, and even dating.[259] Thus, we might anticipate that the individual with an adverse algorithmic negligence score might also be vetted more stringently for a driver's license (if not denied a license altogether), subject to enhanced police scrutiny, and otherwise carefully observed for anticipated careless behavior. The same pattern holds for other status predictions; we would expect similar effects for persons whose score predicted a need for social services, for child-protective observation, or other anticipated misbehavior. What we know about the social operation of algorithmic scoring suggests that even if personalized behavioral-valuation scores did not begin as an explicit social differential, they would rapidly become such.

### A. *Perfidious Transparency*

We have now established that the bias in algorithmic systems lies not so much in their fidelity or infidelity to some objective state of the world, but rather in the feedback loop constituting social shaping of algorithmic input and the corresponding shaping of social perceptions by algorithmic output. Much of the critical literature to date addressing commensurate algorithmic practice has called for greater transparency: for open examination of the data or the mechanics involved in automated decision-making systems.[260] This argument, as we have seen above, is aimed at the problem of governance of algorithms,[261] and is similar to those typically applied to democratizing or inhibiting powerful institutional actors; the underlying supposition is that social damage will be lessened if a governance mechanism's use is acknowledged and its functioning is understood.[262] On this view, biases, omissions, and disparate representation could perhaps be corrected if exposed and identified in the data or data processing of automated systems.

Despite their salutary prospects, the feasibility of such transparency proposals is doubtful, as the technical and social positioning of algorithmic

259 *See* Rona-Tas, *supra* note 155, at 52–53.

260 *See* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 107–08 (2018); Citron & Pasquale, *supra* note 100, at 8, 10–11; Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. Rev. 54, 58–59 (2019); Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1568.

261 *See supra* notes 25–30 and accompanying text.

262 *See* Coglianese & Lehr, *Transparency*, *supra* note 2, at 3.

processes obstinately resists such transparency.[263] First, there are powerful market incentives for the owners or employers of such algorithms to keep the workings of their system proprietary and confidential, in order to maintain market advantage.[264] The design of algorithmic systems, the training data and methods of training, and the databases on which the systems operate are all likely to constitute valuable trade secrets, entailing substantial investment of time, capital, and resources, but easily copied by competitors if disclosed. Thus, private entities in particular are likely to resist requirements for disclosure.

However, a subsequent, second layer of impediment is likely to be more important: even when disclosure is conceded or mandated, the disclosed code is unlikely to be understood by legislators, judges, and policymakers, let alone the average citizen. Disclosure of incomprehensible technical details will not empower either the average citizen or institutional decisionmakers to make autonomous choices about algorithmic systems. Rather, those empowered to make social choices regarding such algorithms must likely rely on the understanding communicated by experts who are in a position to make first-hand assessment of the disclosed algorithmic details.[265] This situation is not altogether unusual where decisions about technical subjects must be made, but unquestionably imposes additional costs, delays, and impediments that may negate many benefits of disclosure.

Finally and most significantly is the problem of algorithmic processing inscrutability: as described above, metrics generated via machine learning are typically based on functions developed without explicit human direction or oversight.[266] Even if the process of data extraction and evaluation is disclosed, and even if experts are available to interpret it for lay decisionmakers, those experts will generally not know why automated systems develop the abductive rules that they do.[267] As described above, machine learning systems develop their own criteria and determine their own results. This is generally considered to be an advantage of the system, but the disadvantage is that it is frequently impossible to understand exactly how the system reached the particular output it did, even if its functional mechanism is understood.[268] Some version of the first two impediments attend almost any new technological development, but this third impediment in particular offers

---

263 Burrell, *supra* note 242, at 10; *see also* Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 978, 981–83 (2018) (explaining why transparency is inadequate for oversight of algorithms).

264 Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 658 (2017).

265 Anton Vedder & Laurens Naudts, *Accountability for the Use of Algorithms in a Big Data Environment*, 31 INT'L REV. L. COMPUTS. & TECH. 206, 208 (2017).

266 *See supra* notes 41–55 and accompanying text.

267 *See* Burrell, *supra* note 242, at 5; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1094 (2018).

268 *See* Mackenzie, *supra* note 1, at 436 (noting that rendering predictive processes visible is a central challenge for machine learning).

significant reasons for skepticism about any degree of meaningful algorithmic transparency.[269]

But set such impediments aside for the moment, recalling that the purpose here is to interrogate the problem of algorithmic governmentality.[270] Assuming that transparency is even achievable, my analysis here raises a different set of questions, indicating that the proffered solution of algorithmic transparency may lend itself to highly problematic, even negative social effects. Transparency is touted as the palliative for the failure of algorithmic accuracy.[271] But algorithmic transparency can be conceived as *compromising* scoring accuracy—if the subjects of algorithmic scrutiny are aware of the factors considered by the algorithm, they may be able to alter the outcome of the scrutiny in predictable ways, so as to manipulate or "game" the outcome.[272] The algorithmic score may then be viewed as inaccurate or prejudiced by the subject's deliberate alteration of data inputs.[273]

As a consequence of the potential for such algorithmic "gaming," some commentators have cautioned against transparency of algorithmic processes, at least in some instances and perhaps in all, warning that consumers who are aware of the parameters for algorithmic scoring might alter their behavior to "game" the system and produce desirable metrics.[274] In some instances an advantageous manipulation might constitute a favorable algorithmic assessment; in other cases an advantageous manipulation might constitute avoidance of the algorithm altogether.[275] The parade of algorithmic gaming horribles that are recited to support this view is led by potential manipulation or evasion of algorithms intended to monitor safety, national security, or other ostensibly compelling protective purposes.[276]

Other commentators in response have discounted the likelihood of such consumer gaming, arguing that even if consumers know the components of an algorithmic output, advantageously affecting those outcomes is difficult, so that the benefits of transparency outweigh the occasional manipulation of

269 *Cf.* Cynthia Dwork & Deirdre K. Mulligan, *It's Not Privacy, and It's Not Fair*, 66 STAN. L. REV. ONLINE 35, 37 (2013) (observing that the complexity of algorithmic systems will preclude transparency alone from eliminating bias).

270 *See supra* notes 25–30 and accompanying text.

271 *See, e.g.*, Coglianese & Lehr, *Transparency*, *supra* note 2, at 16 (discussing tradeoffs between transparency and accuracy).

272 Citron & Pasquale, *supra* note 100, at 30; Kroll et al., *supra* note 264, at 658.

273 *See* Kroll et al., *supra* note 264, at 658; Zarsky, *supra* note 260, at 1554.

274 Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 22 (2018); Kroll et al., *supra* note 264, at 657–58.

275 The gaming of algorithmic metrics thus partakes of what Leo Katz dubs "avoision," a legal and moral gray area between legitimate avoidance and illegitimate evasion. *See* LEO KATZ, ILL-GOTTEN GAINS: EVASION, BLACKMAIL, FRAUD, AND KINDRED PUZZLES OF THE LAW, at ix–x (1996).

276 *See* Bambauer & Zarsky, *supra* note 274, at 33; Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 43 (2013); Zarsky, *supra* note 260, at 1554.

algorithmic outcomes.[277] But each of these views overlooks the effects described here. Both the gaming concern and its response are couched primarily in terms of algorithmic accuracy, that is to say, they revolve around the ability or inability of the measured subject to skew the algorithm away from an accurate metric. This largely misses the critical characteristics of algorithmic scoring as deployed in the private sector, and as practiced in the public sector, which is to strategically reshape the meaning of metric produced.

Consequently, my concern here is the alternative and contrasting view that such algorithmic manipulation might be advanced as a *useful feature* rather than as a flaw. Allowing or encouraging the surveilled subject to rework the factors affecting algorithmic output is a strategy that can be represented as a tool for real or perceived personal achievement. Knowing the factors that are calculated in the algorithm allows the subject to deliberately alter those factors, thus producing a more favorable score. To the extent one believes that the algorithmic output reflects actual status, improving the score may be correlated with personal improvement. The paradigm business model for such personalized algorithmic discipline is the Fitbit device, by which users voluntarily submit to digital monitoring in order to track and improve algorithmic scores regarding their health and physical fitness.[278] But other examples of deliberate encouragement of score alternation, such as "improvements" in personal credit scores, or in *U.S. News and World Report* law school rankings, are plentiful.

Some commentators have excluded such algorithmic self-discipline from the problem of "gaming," naively assuming that such self-improvement tracking poses no threat because it confers social benefits, such as better health, or better personal finances, that result from greater scoring accuracy. This view ignores the social construction of the algorithmic process and its output, assuming that algorithmic scoring in fact reflects some cognate state in the real world. But whether or not this assumption is correct in any given instance, all such personalized behavioral conformation constitutes forms of panoptic self-discipline, altering personal behaviors due to algorithmic surveillance. And in examples such as the Fitbit device, we also unquestionably see instances of market discipline, in which consumers are induced to rely upon and change their behavior according to a particular algorithmic business model.[279]

These two approaches, of algorithmic gaming as inaccuracy and algorithmic gaming as a goal, entail radically different underlying assumptions regarding algorithmic quantification. The position that proper algorithmic metrics are threatened by reactive effects assumes that what is being measured by the algorithm is a static or inherent characteristic in the world. The alternate approach assumes that the algorithm measures a

277 Ignacio N. Cofone & Katherine J. Strandburg, Strategic Games and Algorithmic Transparency (2020) (unpublished manuscript) (on file with author).

278 *See* Esmonde & Jette, *supra* note 217, at 300.

279 *See supra* notes 217–20 and accompanying text.

changeable or dynamic property such that the measurement may be properly harnessed to change behavior.[280] This is a typical tension within socially reflexive, reactive practices: the conception of a given measurement as a neutral and inherent representation of the world, and the conception of that measurement as a vehicle for monitoring accountability and effecting desired performance.[281]

We can discern each of these positions, and their effects, in the example of American credit scoring. We have already said something about the biases and social stratification pervasive in the practice of credit scoring.[282] Transparency of credit scoring criteria and credit scoring practices might be thought to offer a solution to combat such biases, or at a minimum to identify and understand them.[283] The initial position of the financial services industry regarding such transparency as to the algorithm was that disclosure of the scoring criteria would compromise accuracy by allowing the subjects of the assessment to manipulate their own scores.[284] Later, this position changed, and a measure of credit scoring transparency was embraced as a means to allow consumers to deliberately engage in market behaviors that would elevate their credit scores.[285] The altered consumer behavior was accepted as a way to achieve better credit access, and by improving consumers' histories of financial activity, allow them to gain enhanced access to financial products.[286]

It is worth noting that the shift in conception of credit scoring is unlikely to have occurred because of a sudden enlightenment or concern for consumers, either on the part of credit scoring agencies or on the part of their commercial-lending-institution subscribers, regarding the autonomy of the scored populace. One rather suspects that the change occurred in order to expand the market, either for the financial products of credit scoring subscribers, or perhaps the market for credit scores themselves. Allowing the subjects of credit scoring to improve their scores provides a greater number of customers eligible for the financial products on which the loans depend. And of course, in order to improve their scores, the subjects of scoring need to know and self-monitor those scores, which are typically provided for a fee.

This should make clear that transparency of algorithms may be a strategic move on the part of algorithmic decisionmakers, facilitating the deliberate employment of commensurate and self-fulfilling social processes to the advantage of the decisionmaker. Some previous commentary has recognized

280 Espeland & Sauder, *supra* note 128, at 7.

281 *Id.*

282 *See supra* notes 100–09 and accompanying text.

283 *See* Citron & Pasquale, *supra* note 100, at 24 (advocating transparency as a curative to credit scoring inequities).

284 *Cf.* Kear, *supra* note 100, at 361.

285 *See* Fourcade & Healy, *supra* note 100, at 565; *see also* Donncha Marron, *Debt, Consumption and Freedom: Social Scientific Representations of Consumer Credit in Anglo-America*, HIST. HUM. SCIS., Oct. 2005, at 25–26 (describing the shift in social perception of personal debt from a moral problem to a financial tool to an individual entitlement).

286 Kear, *supra* note 100, at 361.

that decisionmakers may behave strategically by hiding algorithmic bias in order to avoid accountability, or might disclose algorithmic features in order to deter gaming by subjects.[287] These concerns again relate to algorithmic accuracy. But I have shown here that disclosure may represent deliberate behavior in a "long game" by those deploying the algorithm: not to avoid responsibility for poor accuracy, nor to discourage input gaming by those subjected to algorithmic surveillance, but rather to purposely manipulate the behavior of algorithmic subjects toward some desired surveillant disciplinary outcome. Indeed, careful analysts of the new "data-driven" economy have observed precisely this type of disciplinary strategy.[288]

Algorithmic discipline is not merely a business strategy, although it may be exploited as such. To the extent that transparency in predictive algorithms is feasible, transparency as to inputs facilitates self-direction of algorithmic scoring, determining whether the score is positioned as dynamic or static, autonomous or destined, aspirational or conditional, normative or descriptive. Bias figures into the picture not as a matter of decisional accuracy, but primarily as it serves to create differential opportunities for the wealthy and the impoverished to deliberately shape their calculated "data doubles" by engaging in recorded behaviors intended to affect algorithmic scoring.[289] And to the extent that these effects may be exploited in the market, algorithmic transparency may actually facilitate such exploitation.

As a consequence, calls for algorithmic transparency in order to improve accuracy are to some extent orthogonal to the reflexive effects that I have outlined here; we must expect reflexivity costs to occur whether or not algorithms are transparent. But far from objectively describing or predicting social behavior, or even simply influencing social outcomes, algorithmic scoring constitutes a tool for deliberately manipulating self-determination. Transparency of algorithmic inputs results in algorithmic outputs such as the credit score, not to mention metrics such as the self-surveillant Fitbit output, becoming situated as dynamic self-improvement projects reflecting the moral fiber of the individual.[290] When considered in the context of algorithmic legal scoring, transparency becomes a normative and moral consideration, and not necessarily in a beneficial way.[291]

### *B. Quantified Performance*

The reflexive biases described here lead inexorably to an additional set of issues arising from the proposed use of algorithmic scoring to determine legal status: the implications of such scoring for normative or moral expecta-

---

287 *See* Cofone & Strandburg, *supra* note 277, at 3–5.

288 *See* Cohen, *supra* note 215, at 148, 151; Marion Fourcade, *The Fly and the Cookie: Alignment and Unhingement in 21st-Century Capitalism*, 15 SOCIO-ECON. REV. 661, 672 (2017); Zuboff, *supra* note 214, at 75.

289 Kear, *supra* note 100, at 363.

290 MARRON, supra note 258, at 193.

291 *See* Citron & Pasquale, *supra* note 100, at 18.

tions.[292] These issues proceed, again, directly from the known social and cognitive dynamics of quantification, but have particular importance for algorithmic calculations situated in a legal determination. Quantification, such as that undertaken by an algorithmic scoring system, always implies sorting or categorization.[293] Algorithmically determining an individual's legal character, such as negligence or reasonable prudence—even assuming a credible determination were possible—would by definition require measurement and comparison of the consumer's predicted characteristics or behaviors against the those of other individuals. The entire project assumes that the profiled subject, and those against whom the subject is compared or from whom the subject is distinguished, fit into stable categories. As described above, machine learning assumes the presence of stable and distinct categories that can be detected by pattern analysis; the system learns to discern distinctions among the features of such fixed categories.[294]

Such an assignment of classification or categorization, although couched in seemingly objective numerical representation, is never a neutral or endogenous act; it is rather a value-laden imposition of a particular social order on its subjects.[295] Such values are, as we have seen, embedded in the processes that produce numerical outputs.[296] This is in part the concern articulated by many commentators regarding covert bias in the processing of surveillance data.[297] But regularized measurements also acquire new meanings derived from the social roles that they inevitably take on. Such infusion of value judgments into metrics proceeds, again, by means of well-known social mechanisms. Imposition of any type of category, with its accompanying categorical standard, necessarily elevates some viewpoints and suppresses others.[298] Items that resemble one another may be categorized together, but what it means to "resemble" one another is the outcome of a contested social process.[299] Indeed, the process may well be inverted, as items that are classified together become socially perceived as similar, and items that are classified apart from another become socially perceived as different.[300] Such percep-

292 In a recent commentary paralleling my critique here, Philip Bender has relatedly questioned the normative gaps in personalized law conjectures. *See* Bender, *supra* note 19, at 2.

293 Fourcade & Healy, *supra* note 42, at 289.

294 Mackenzie, *supra* note 1, at 433.

295 *See supra* notes 83–91 and accompanying text.

296 *See supra* notes 92–99 and accompanying text.

297 *See supra* notes 237–39 and accompanying text.

298 GEOFFREY C. BOWKER & SUSAN LEIGH STAR, SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES 5 (1999).

299 *See* HARRIET RITVO, THE PLATYPUS AND THE MERMAID AND OTHER FIGMENTS OF THE CLASSIFYING IMAGINATION, at xii–xiii (1997).

300 *See* ÉMILE DURKHEIM & MARCEL MAUSS, PRIMITIVE CLASSIFICATION 81–82 (Rodney Needham ed. & trans., 1963).

tions eventually crystalize as "social facts" that perpetuate themselves and undergird broad networks of normative belief and behavior.[301]

Quantified categorization is a frequent social process, and the imposition of categories may occur pursuant to a number of different classificatory purposes. But even when directed to different ends, categories are volatile and to some degree interchangeable. To distinguish their epistemic and social functions, Marion Fourcade divides categorical ranking systems into three broad purposive goals: first, those rankings that are nominal, categorizing their subjects according to the characteristics of the item ranked; second, those that are cardinal, which categorize according to quantified metrics; and third, those that are ordinal, which rank objects relative to one another.[302] These three systems rely on ostensibly different criteria, but despite their operational differences, there is considerable slippage at the intersection of these systemic types.

For example, ordinal rankings are relative rankings, but such rankings typically imply valuations by which subjects, including persons, are ranked relative to one another: performance, intelligence, quality, efficiency, and so on. Relative ordinal rankings thus imply nominal categorization according to subject characteristics and slip into becoming effective nominal rankings. Ordinal rankings also quickly become quantified into numerical hierarchies, and so morph into ostensibly cardinal rankings. And, just to complete the circle, the metrics associated with cardinal rankings frequently become perceived as characteristics of the items ranked, and so blur into a nominal ranking.

Thus, to illustrate, if I award a particular student the highest grade in the class, that cardinal ranking quickly slips into a relative ordinal ranking with respect to the other students—the ordinal ranking may be granular, such as ordering the students as first, second, third, and so on, or it may be more clustered, such as ordering the "A" students, the "B" students, and so on. Probably the ordinal ranking will be based on some cardinal ranking, such as grade point average or class "participation points." To the extent that either the ordinal or cardinal ranking is attributed to superior intelligence or effort, it rapidly becomes a nominal ranking for the "best" or "brightest" or "smartest" student in the class. And of course the lowest rankings quickly become nominalized as well, labeling such students as the dullest or laziest.

These effects are socially pervasive and can be expected to manifest themselves in proposed legal metrics, where slippage between different types of rankings is likely to occur due to the embedded assumptions undergirding assigned legal status. Consider again the example of personalized negligence prediction and the conceptual baggage that comes along with it. Negligence

301 Emile Durkheim, The Rules of Sociological Method 50–51, 74 (Steven Lukes ed., W.D. Halls trans., 1982); *see also* Alain Badiou, Number and Numbers 2–3 (Robin Mackay trans., 2008) (observing that "whatever produces number can be culturally located").

302 Marion Fourcade, *Ordinalization: Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014*, 34 Socio. Theory 175, 176–78 (2016).

is in the current era most often conceptualized in terms of utility, framed as economic efficiency.[303] That framework is the dogma associated with the application of the famous "Learned Hand inequality" from *United States v. Carroll Towing*, which has become an iconic algebraic expression of optimization for the costs of spending on safety against the costs of accidents.[304] Within that framework, the market behavior of individuals, such as a given individual's propensity to carelessness, or willingness to pay for safety precautions, is typically treated as a nominal or essential quality that serves to order private exchanges as discussed above with regard to rationality and preferences.[305]

Under the neoclassical economic tort models on which negligence predictive proposals are ultimately based, an individual's willingness to invest time or resources in a given behavior is treated as an endogenous characteristic of that individual, and as such may be considered as an intrinsic nominal categorization. But to the extent that this categorization arises out of the numerical mastication of data profiles by an algorithm, it will be treated as a cardinal category, as simply an ordering of consumers along a demand curve as reflected by their assigned metrics.[306] The quantification of the ranking may, however, additionally be seen as comparative, ordinally ranking one consumer relative to another. And such relative technological ranking of the individual lends itself ultimately to nominality, in which the algorithmic score becomes an indicator of self-management and "reasonably prudent" behavior, reflecting consumers' preferences and safety investment as manifest in the ordinal algorithmic score.[307]

Thus, categorical cognitive slippage across classifications means that there will inevitably be attribution of moral or normative meaning to algorithmic scoring. Because the algorithmic score, whether purporting to represent creditworthiness, intelligence, popularity, or trustworthiness, appears objective, any positive or negative change will be attributed to the character or actions of the individual being scored rather than to bias, feed-

---

303 The literature characterizing tort law in this fashion is enormous; the most germinal work is likely GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 24, 26–29 (1970). *See also* WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 85–89 (1987) (explaining negligence law in terms of economic efficiency).

304 *See* United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947) (expressing the conditions for tort liability as occurring where the burden of safety precautions is less than the magnitude of potential harm times the probability of harm, $B < PL$); *see also* LANDES & POSNER, *supra* note 303, at 85–89 (characterizing negligence law in terms of the Learned Hand inequality); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 163–66 (4th ed. 1992) (same).

305 *See supra* notes 172–77 and accompanying text.

306 *See* Christopher Townley, Eric Morrison & Karen Yeung, *Big Data and Personalized Price Discrimination in EU Competition Law*, 36 Y.B. EUR. L. 683, 684 (2017).

307 *See* Fourcade, *supra* note 302, at 188–89; *see also* Lury & Day, *supra* note 77, at 29 (discussing the translation of ordinal ranking into personal optimization).

back, or variance in the algorithm.[308] Algorithmic metrics appear to objectively reflect chosen behavior, and so become an ethically significant indicator of individual character.[309] Algorithmic calculation thus helps precipitate a shift from *measurement* to *judgment*. The algorithmic result comes to quantify individual performance, which is recorded and represented as a stream of autonomous choices.[310] Credit scores, fitness monitoring, and other algorithmic metrics come to imply the responsibility to control one's own providence, using algorithmic quantification as the measure of self-determination.[311]

These effects are accelerated by the deployment of algorithms intended for self-surveillance and algorithmic discipline, as illustrated by the examples above.[312] And as should also be clear from the negligence example, coupling algorithmic quantification with the hollow premises of neoclassical economics reinforces this effect. We have seen that the ostensibly neutral assumptions of economic models are not blank slates, but instead are rapidly imbued with social meaning.[313] Once the assumption of plentiful, even perfect information supplied by Big Data is accepted, then under neoclassical models, behavior can no longer be ascribed to bounded rationality or imperfect information.[314] Rather, any behavior must be considered to constitute informed choices reflecting manifest preferences, and the outcomes of the behaviors are deserved. Outcomes then become viewed as morally chosen and deserved consequences.[315] Algorithmically quantified behaviors become viewed as indicators of autonomous decisions, individual merit, and social worth.

Thus, algorithmic scoring delivers something besides a neutral numerical metric. Returning again to the example of economic categorization, metrics such as the American credit score entail a moral dimension, assigning to the recipient an implicit level of trustworthiness, self-discipline, and social status that affects the recipient's economic and social opportunities, including employment, investment, and dating opportunities.[316] To the extent that these social valuations are considered to be based on measurements, they may be ascribed to ostensibly neutral cardinal rankings.[317] But to the

---

308 Fourcade & Healy, *supra* note 41, at 25.

309 *Id.* at 24.

310 Fourcade & Healy, *supra* note 100, at 565–66.

311 Fourcade & Healy, *supra* note 41, at 24.

312 *See supra* notes 229–36 and accompanying text.

313 *See supra* notes 170–81 and accompanying text.

314 Fourcade & Healy, *supra* note 41, at 24.

315 *Id.*

316 *See, e.g.*, Michelle Singletary, *When It Comes to Dating, Where Might Bigger Really Be Better? Credit Scores*, WASH. POST, May 14, 2017, at G04; *see also* Jane Dokko, Geng Li & Jessica Hayes, *Credit Scores and Committed Relationships* 1 (Fed. Rsrv. Sys. Fin. and Econ. Discussion Series, Working Paper No. 081, 2015), https://www.federalreserve.gov/econresdata/feds/2015/files/2015081pap.pdf (exploring positive correlation between high credit scores and relationship longevity).

317 Fourcade, *supra* note 302, at 178.

extent that valuations are perceived as based on characteristics of the subject, they may be ultimately understood as objective or natural *nominal* rankings.[318] We may expect that an algorithmic categorization for legal metric such as negligence, culpability, or reasonableness will similarly collapse under its own weight into a comparable moral singularity, even if the measurement is ostensibly neutral. And consequently, using such scores to affect legal status implicates the moral framework underlying such status.

## V. Calculated Culpability

We turn then finally from normative judgments to legal judgments, to considering the intersection between algorithmic legal metrics and the moral framework undergirding legal decisionmaking. The known progression of cognitive slippage from quantification toward evaluation, expectation, and judgment has profound implications for both the proliferation of quantifying algorithmic systems and for proposed responses to such proliferation. If, as Fourcade and Healy argue, social processes transform algorithmic metrics into "moral injunctions,"[319] then the use of algorithmic legal metrics must unavoidably alter the construction of juridical injunctions. Legal metrics provide a cardinal categorization of their subjects, but this will slip into ordinal, comparative categorization of their subjects. As this characterization in turn slips toward nominal categorization, becoming a moral characterization or judgment of the subject, it infects the moral or normative judgments made by legal decisionmakers.

The negligence liability example used above illustrates this intersection; we have already seen how algorithmic negligence metrics will eventually slide toward nominal categorization, implicating the inherent responsibility or irresponsibility of the subject.[320] Yet because it is inextricably linked to *ordinal* or comparative ranking among citizens, the assumed preference for safety or risk spills into the public sphere, comparing the individual's quantified prudence to the normative expectations of other individuals.[321] Negligence tort liability in particular is intended as private ordering in the service of public order, engaging privately initiated legal claims to redistribute the social costs of harmful activity. But categorization slippage means that reasonably prudent behavior will never be merely viewed as a matter of social efficiency, but rather also as a matter of personal qualities and accountability.

The slippage in treatment of legal metrics thus invokes a dichotomy fundamental to democratic society: liberal democracies simultaneously assert that the autonomy and individual worth of the person are to be respected, while at the same time asserting that no one is entitled to privileged treat-

318 *Id.* at 179.

319 Fourcade & Healy, *supra* note 41, at 24.

320 *See supra* notes 303–310 and accompanying text.

321 *See* Fourcade, *supra* note 302, at 180.

ment under the law.[322] Democratic regulation becomes illegitimate when it is arbitrary, but may be equally illegitimate when it is impersonal and calculated—too much personal variation violates democratic principles of equality, but too little violates liberal principles of autonomy.[323] An "unjust" legal regime may mean a regime that undervalues either the former or the latter.

To maintain democratic legitimacy, the imposition of a "personalized" legal category must therefore accommodate both autonomy and social equality.[324] From this standpoint, legal liability based on a nominally characterized "personalized" or "granular" negligence score may seem to depart from values of equality under which actors are all subject to uniform law. At the same time, ordinal treatment of the score may seem to depart from values of autonomy, under which the individual is autonomously valued. Significantly, the judgment of liability quite separately carries with it a strong deontological sense of culpability or blameworthiness. As Herbert Morris observed in the criminal context, a finding of guilt or liability is in some sense a backhanded recognition of a defendant's autonomy—this is denied to minor children or those lacking autonomous mental capacity, who are deemed unable to exercise the self-determination necessary to enter into contracts or to commit crimes.[325]

This raises a familiar question as to the purposes of legal liability: whether a judgment of such legal status constitutes a statement of blameworthiness or merely a utilitarian tool of deterrence. The former role for legal determination assumes a particular type of moral responsibility and implies that the subject could and should have made different choices, might feel remorse, might make better future choices, might improve himself. The latter role for legal determination does not necessarily preclude any of those concepts, but is largely indifferent to guilt so long as infringement is deterred, either specifically or generally. On an economic version of utilitarianism, the imposition of a particular judgment should create and reinforce incentives that will most efficiently distribute social resources.

The issue of culpability may be especially problematic when the decision regarding the defendant's liability is based on an algorithmically scored metric. I have shown above that algorithmic scoring is anything but neutral or objective, even though the bland numerical appearance of a score lends itself to objective characterization.[326] However, the characteristics of algorithmic scoring that I have reviewed above—including, ironically, their ostensible neutrality—tend toward moral inferences regarding the autonomy of the scored individual.[327] There is considerable evidence that sorting and classifi-

322 *See* Andrew Abbott, *The Future of the Social Sciences: Between Empiricism and Normativity*, 71 Annales HSS 343, 352 (2016); Wendy Brown, *Wounded Attachments*, 21 Pol. Theory 390, 400 (1993).

323 *See* Bender, *supra* note 19, at 62.

324 *See id.* at 54, 62.

325 *See* Herbert Morris, *Persons and Punishment*, 52 Monist 475, 485, 493 (1968).

326 *See supra* notes 92–99 and accompanying text.

327 *See supra* notes 306–09 and accompanying text.

cation schemes never confine themselves to objective application, no matter how neutral their intent.[328] Rather, the nature of classification and ranking systems lends itself to moral labelling, and such labelling is a commonplace and well-studied characteristic of the emerging algorithmic society.[329]

We should therefore fully expect that this type of labeling will occur should we embark on negligence or other liability rankings, especially since not only algorithmic judgments, but a system of legal judgment, would have been engaged in such a process. Indeed, we should expect that the moral valence of the legal judgment will be quickly infected by the imputed morality of quantification, described above.[330] To the extent that a judgment of legal liability entails a finding of blameworthiness, tying such liability to a standard generated by quantified behavior poses both practical and jurisprudential conundra: Is the "personalized" metric on which legal status would be based intended as a representation of the consumer's intrinsic character or as a target for self-surveillant adjustment?

We have seen that either view may be sustained by algorithmic scoring. The score itself, although value-laden, does not dictate the purposes to which those values are put when introduced into legal processes. Take again the example of the algorithmically personalized negligence score. Assuming that we treat the score's metric of expected carelessness as a stable and inherent quantity, what policy should guide the assignment of legal status arising out of the algorithmic metric? For example, should a score indicating a lack of prudence mean that the subject's negligence liability should be ratcheted up, in order to increase that person's incentives to take care, so as to bring him closer to the average social expectation for safety? Or should the standard in fact be "personalized" in light of the score, in effect lowering the expectation for prudent behavior, on the theory that the person should simply take the degree of care that *he* reasonably can, given his inherent disability for prudence?

One might of course fabricate the pretense of an answer to this dichotomy by draping it in the rhetoric of costs and benefits; advocates of personalized negligence law have to some extent done exactly this.[331] But set aside the elaborate ruminations of economic analysts as to whether tort law should or should not encourage improvement of those truly prone to negligence, because their improvement might inefficiently cost more than the value of gains to safety.[332] Consider instead the normative expectations embedded in the existing legal system and the known and knowable practices of the society in which that system operates. Either of the diametrically opposite outcomes that I have identified is doctrinally plausible: the former approach effectuates the liberal value of equality; the latter approach effectuates the liberal value of individuality.

328 *See supra* notes 292–301 and accompanying text.
329 *See supra* notes 305–09 and accompanying text.
330 *See supra* notes 313–19 and accompanying text.
331 *See, e.g.*, Ben-Shahar & Porat, *supra* note 13, at 655.
332 *See id.* at 652–655.

And, given what I have shown here regarding the portability of numerical commensurates, and the known social slippage of other algorithmic metrics such as credit scores, we might imagine that an adverse negligence score will migrate to other settings, taken as an indicator that the individual is literally an accident waiting to happen. Certainly we should expect that it would quickly become the subject of pretrial discovery and evidentiary consideration as an indicator of the defendant's propensity toward negligence. It would not be implausible to anticipate that a sufficiently adverse score could become the trigger for some type of presumption or evidentiary burden-shifting along the lines of tort law's current res ipsa loquitur doctrine, on the theory that "the score speaks for itself."

Assume now that we do not treat the score as measuring a static or constant property, and instead adopt the view that whatever measurable tendency toward carelessness an individual may possess is not an immutable quality, but subject to personalized improvement. If an algorithmic legal score is considered dynamic rather than static, subject to alteration by the person represented, in the manner of improved credit scoring, or "quantified self" calculative development, then consumers could presumably work to achieve a "better" score. As we have seen, this type of shift trades a view of static, endogenous character for a cybernetic view that elevates assumptions of autonomy—and here I use "cybernetic" in its original context, to designate a system of control or governance.[333]

But it is unclear what this would mean for quantified individuals caught in the dilemma of algorithmic measurement. For example, in constructing a personalized duty of care, is the individual permitted to know the algorithmic methods and factors that go into his negligence score? Would doing so "compromise" the tort score, or is the individual afforded the chance to improve his algorithmic negligence prediction?[334] If we view the score as a changeable metric, might we expect the emergence of a class of "prudent care coaches," devoted to helping potential tortfeasors elevate their scores?[335] To the extent that the tort system is intended to foster prudent and careful behavior, would such dynamic scoring achieve a goal of the system, or might it prompt inefficient overinvestment in score manipulation rather than efficient spending on safety?

These concerns are by no means limited to the problem of negligence scoring, but rather present themselves wherever algorithmic scoring and legal status meet. Consider the example raised above of the individual who is algorithmically identified as a proper target for social welfare services.[336]

333 *See* Norbert Wiener, Cybernetics: Or Control and Communication in the Animal and the Machine 11–12 (2d ed. 1961); *see also* Yeung, *supra* note 3, at 507 (noting that the roots of algorithmic regulation lie in cybernetics).

334 *Cf. supra* notes 282–84 and accompanying text (discussing the now-defunct concern that transparent credits scoring could be manipulated).

335 *Cf. supra* notes 138–39 and accompanying text (discussing the emergence of credit score repair businesses).

336 *See supra* note 8 and accompanying text.

The provision of public assistance has perennially been associated with a morass of social judgments regarding self-sufficiency, personal competence, responsibility, and autonomy. The predictive score generated for such a person might be taken as a measurement of the individual's extrinsic circumstances indicating a need for assistance—or it might be taken as a measurement of the individual's intrinsic character indicating a need for assistance. In either instance, the intrinsic biases and distortions of the metric become the pattern for the self, certified by the public administrative system that generates and adopts the score.

Thus, would a "better" or more desirable welfare score be the score that qualifies the subject for public assistance and welfare services, or would it be the score that allows the subject to avoid qualifying for such assistance? A qualifying score would ensure the availability of assistance that might be desperately needed, but might also be perceived as a badge of personal failure or irresponsibility. Should the criterion for such scoring be made "transparent" to facilitate accuracy and governmental oversight, or would this allow recipients to "game" the distribution of services? Or should the algorithmic metric be made transparent so as to be used as a "self-improvement" metric to guide or encourage welfare recipients toward self-sufficiency? Might a transparent score be manipulated instead to embezzle public services—or more likely, might it be *perceived* by fiscal watchdogs as being used in such nefarious ways?

Consider as well a third example from current algorithmic governance: Should the parent whose children are algorithmically deemed at risk be transparently informed of the criteria and parameters for her administrative profile?[337] Should she be warned of the threshold calculation that will trigger state intervention to remove the children from danger? Would doing so perhaps allow her to "game" the calculation so as to avoid the intervention, placing the children at greater, surreptitious risk? Or would disclosure, alternatively, provide a metric for self-improvement, quantifying her risky behaviors so as to prompt mitigation and conscious minimization of those behaviors? Should she perhaps be coached by social workers or private advisors on how to obtain a better "abuse score" through behavioral modification? Determining such status on the basis of an algorithmic metric effectively entails a policy choice between autonomy and determinism, shifting the individual's performance of self *and* legal institutional endorsement of that performance.

We are then left to wonder: What incentives would be created by reliance on algorithmic scoring? Our understanding of human-algorithm interactions tells us that algorithmic scoring of the tort defendant, the welfare recipient, or the troubled parent would mark a change in agency, from quantified subject to quantified object.[338] Algorithmically determined legal scoring might be intended to quantify a static and intrinsic metric. But common

337 *See supra* note 5 and accompanying text.

338 *See* Espeland & Sauder, *supra* note 128, at 7.

experience tells us that human character is not static and that preferences change, and we now understand that they will do so in part because of reactivity to the algorithms themselves. This does not merely present the problem of picking a particular point in time to assess a given data double, as tricky as that determination might be. A more serious anomaly is that the classifications into which data doubles are slotted are dynamic, and the subjects of legal metrics will begin to perform the classifications assigned by the algorithms. As Jonathan Simon once noted, such actuarial practices substitute governance of moral agents for predictive analysis of risk, altering the perception of self and community.[339]

Reflexive effects will occur on both sides of the evaluation, and indeed within the structure of legal institutions themselves.[340] Thus, although algorithmic legal metrics might be intended to simply assist or enhance the judgmental capability of the existing legal order, what we know of their operation indicates that they inevitably alter or entirely displace the normative structures they are intended to quantify. In either case the criteria and operation of the algorithm, as well as its mere presence in the network of legal metric functions, serves to restructure decisionmaking and to restructure its subjects into calculated citizens, to respond to their own quantification. Deployment of algorithmic legal metrics threatens to supplant the underlying assumptions of law, norms, and reason of neoliberal governance with recursive behavioral protocols and performative algorithmic functions.[341]

## Conclusion

The job of unpacking the intricate, concatenated effects of algorithmic profiling is a complex and daunting task. In this Article I have traced one set of troublesome strands in the tangled skein: algorithmic processes, like economic modeling, engage reflexive social processes to create their own social facts. When deployed in combination, these numerical modeling systems reinforce one another to induce disciplinary behaviors in the social worlds where they operate. This will occur regardless of intent, but is increasingly the basis for purposive restructuring of markets and market participants in the service of what Shoshana Zuboff has termed "surveillance capitalism."[342] When deployed in a legal setting, we should similarly expect algorithmic discipline to intersect with the normative judgments of the legal system to produce perverse outcomes that will not be ameliorated by the most liberal application of transparency, due process, or similar institutional palliatives.

In some senses, the reflexive outcomes I have detailed here are nothing new; commensuration and related processes are ubiquitous social mechanisms, operating in both automated and conventional social settings. The

339 Simon, *supra* note 116, at 773; *see also* Yeung, *supra* note 18, at 266–68 (articulating risks of algorithmic profiling to communal norms).

340 *See* Burk, *supra* note 26, at 305.

341 Parisi, *supra* note 27, at 94.

342 *See* Zuboff, *supra* note 214, at 75.

corrosive effects associated with algorithmic metrics have always been present in social processes, but these have been previously tolerable in the quantities and frequency experienced at the scale and scope encountered via normal human interaction. The worst effects of past social metrics could even be kept in check by a known institutional mechanism of the sort commentators are now invoking, such as due process, notice, and transparency. Indeed, many of the illustrations I have used, such as university ranking or credit scoring, antedate automated quantification.

But this familiarity makes the absence of any consideration of such effects in the deployment of algorithmic legal metrics all the more shocking. And it makes all the more worrisome the panoply of effects that attend automation of reflexive processes, the features of automated decisionmaking that are indeed new: the scope, reach, and speed of automated systems; their pervasive illusion of objective numerosity; the heightened opportunity for slippage between judgment and measurement; the shifting of political and social accountability to the machine; and the potential abdication of value judgments to technical design choices.

Previous commentators have noted some of these features in commercial settings and have expressed the concern that algorithms and related information technology may allow the clandestine manipulation of the users; they worry that such manipulation violates or subsumes autonomy, and so erodes certain fundamentals of liberal democracy.[343] This is a credible concern with extremely serious implications, closely related to the construction of "calculated publics" that I discuss above.[344] But my concern here goes much farther: in the context of legal metrics, algorithmic systems, whether deemed manipulative or not, will inevitably have caustic effects in the context of legal determinations, which entail certain theories about autonomy, and which are subverted by the operation of algorithmic scoring systems.

Absent recognition of such dangers, the prospect of employing predictive algorithms to determine legal status carries a superficial appeal. The promise of anticipating socially harmful behaviors before they occur seems a progressive and alluring aspiration. Similarly, the promise of "personalizing" law so that regulation better fits the circumstances of a given individual seems instinctively an equitable and attractive proposition. Yet when one carefully unpacks the technical, social, and epistemic baggage that such proposals bring to the legal system, it quickly becomes apparent that the plausibility of such proposals is indeed superficial. In particular, such proposals evince little awareness that algorithmic processes, far from providing a magical respite from troublesome information costs, implicate a shift in regulatory commensuration from description to discipline.[345]

Recognizing the deleterious potential of legal metrics does not necessarily preclude any role in democratic society for the use of predictive algorithms. It is one thing for a municipality to use a predictive algorithm to

343 *See, e.g.*, Susser et al., *supra* note 213, at 34–44; Yeung, *supra* note 18, at 267–68.

344 *See supra* notes 218–19 and accompanying text.

345 *See* FOUCAULT, *supra* note 30, at 183–84.

anticipate snarled vehicular traffic so as to place stop lights; it is quite another thing for a municipality to use a predictive algorithm to anticipate criminal behavior so as to place police officers. It is one thing for a hospital to use a predictive algorithm to determine the most effective medical treatment; it is quite another thing for a court to use a predictive algorithm so as to determine liability for failed medical treatments. What I have shown here is that the performative features of legal metrics are a danger to legal institutions. Other applications of predictive algorithms in society will have their own failings, but not necessarily these failings.[346]

Even within democratic governance systems, there is no reasonable hope that, whatever their effects, algorithmic legal metrics will not be relied upon, at least in some situations. As I have shown at the beginning of the Article, such reliance is already well underway.[347] If nothing else, one has to assume that budgetary pressures to accomplish more with fewer resources will drive legal institutions toward the use of algorithmic scoring.[348] Thus, tracing the flawed logic of predictive and personalized law proposals is a crucial undertaking. Algorithmic metrics are socially toxic, but much like biologically toxic substances, they might be useful or even curative if applied in limited and judicious circumstances. At a minimum, we will want to limit exposure to such metrics in order to minimize the harm they will inevitably cause. Understanding the likely impact of their implementation allows us to identify the harms they will inflict if deployed, and to begin identifying situations where such harms are sufficiently repellant that selective curtailment of algorithmic predictions is vital.

346 *See, e.g.*, William Magnuson, *Artificial Financial Intelligence*, 10 HARV. BUS. L. REV. 337, 375 (2020) (arguing that machine learning systems in the finance sector will amplify human error).

347 *See supra* notes 4–11 and accompanying text.

348 *See* Yeung, *supra* note 3, at 514.