Northwestern Journal of Technology and Intellectual Property

Volume 15 | Issue 3 Article 2

Spring 2018

Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias

Anjanette H. Raymond
Kelley School of Business & Maurer School of Law, Indiana University

Emma Arrington Stone Young
Indiana University

Scott J. Shackelford
Indiana University

N O R T H W E S T E R N
JOURNAL OF TECHNOLOGY
AND
INTELLECTUAL PROPERTY

Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias

Anjanette H. Raymond, Emma Arrington Stone Young, & Scott J. Shackelford

April 2018 VOL. 15, NO. 3

Volume 15, Number 3 (2018)
Northwestern Journal of Technology and Intellectual Property

BUILDING A BETTER HAL 9000: ALGORITHMS, THE MARKET, AND THE NEED TO PREVENT THE ENGRAINING OF BIAS

Anjanette H. Raymond,[*] *Emma Arrington Stone Young,*[**] *& Scott J. Shackelford*[***]

ABSTRACT—As sci-fi fans will recall, the movie *2001: A Space Odyssey* is focused on the interaction between humans and artificial intelligence. In the movie, HAL (Heuristically programmed Algorithmic Computer) 9000 computer is an artificial intelligence and the onboard computer on the spaceship *Discovery 1*. HAL 9000, more commonly called "Hal," is capable of many functions, such as speech, facial recognition, lip reading, interpreting emotions, and expressing emotions. HAL is built into the *Discovery 1* spacecraft, and is in charge of maintaining all mechanical and life support systems on board. As the movie progresses, the astronauts become concerned about HAL's behavior and agree to disconnect him, in essence killing HAL. HAL becomes aware of the plan and seeks to stop his death as the movie plot climaxes in a conflict between intelligent machine and his human controllers. Interestingly, *2001: A Space Odyssey* author Arthur C. Clark could not have been more accurate about one of the emerging conflicts to face humanity: what role does society want machines to play in coordinating and governing human activity? The debate resonates from the shared economy to the ethics of artificial intelligence. This article seeks to advance the debate about the need for data regulation that focuses on the *impact* of the *use* of the data. First, it provides a brief explanation of data analytics, algorithms, and machine learning. Second, the article explores some of the common mistakes associated with data modeling within algorithmic processes. Third, the paper explores the impact of the use of data, specifically data that is used to create a digital personhood, to inform algorithms that perform basic services. Fourth and finally, the article seeks

[*] Associate Professor, Department of Business Law and Ethics, Indiana University, Kelley School of Business; Director, Ostrom Workshop Program on Data Management and Information Governance; Adjunct Assistant Professor of Law, Maurer School of Law, Indiana University.

[**] Professional Para-Academic Specializing in Interdisciplinary Discourse & Ethics.

[***] Associate Professor, Indiana University; Research Fellow, Harvard Kennedy School Belfer Center for Science and International Affairs; Director, Ostrom Workshop Program on Cybersecurity and Internet Governance; Senior Fellow, Center for Applied Cybersecurity Research.

to define an ethical decision-making model and regulatory structure for data focusing on the impact of the use of the data upon the individual and society.

INTRODUCTION 216
I. ALGORITHMS, ANALYTICS, AND MACHINE LEARNING 220
II. INVISIBLE BIAS AND THE MECHANISMS OF REPRODUCTION 222
A. The Misuse of Statistics and Social Science 223
B. Engraining of "Common Sense" 226
C. Elided Agency 229
III. SOCIETAL IMPACTS OF BIAS 232
A. Barriers to Economic Development and Basic Life Services 233
B. Power Differentials & Economic Incentives in the Information Marketplace 234
C. Influencing the Information Received 237
D. Surveillance Based Influence and Manipulation 239
IV. UNPACKING THE REGULATORY LANDSCAPE 240
A. A Polycentric Primer 240
B. The Federal Trade Commission 241
C. Global Data Laws and Regulations: Non-Discrimination Case Study 243
D. Non-Discrimination Legal Regulation 244
E. Implications for Policymakers and Managers 248
V. AN ETHICAL MODEL FOR EXAMINATION OF NEW DILEMMAS 250
CONCLUSION 253

INTRODUCTION

The technologies of collection and analysis that fuel "Big Data" are being used in nearly every sector of society and the economy;[1] in fact, data collection in large parts of the developed world is nearly ubiquitous.[2] Unsurprisingly, much of the information gathered has to do with consumers whose information is of high value to businesses seeking to gain and retain

[1] *See, e.g.*, European Commission, *EU-Funded Tool to Help Our Brain Deal With Big Data*, EUROPA PRESS RELEASE (Aug. 11, 2014), http://europa.eu/rapid/press-release_IP-14-916_en.htm (last visited Aug. 20, 2014) (discussing new research funding into Big Data in the European Union) [https://perma.cc/V8LD-HX8E].

[2] *See e.g.,* FEDERAL TRADE COMMISSION, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (May 2014).

customers through tailored advertising and services.[3] As greater value is recognized, more information is collected,[4] and the cycle continues.

Many consider it a benefit that advertisements can be directed and tailored to the individual instead of bulk mails and communications that clog inboxes and mailboxes worldwide.[5] And the use of algorithms combined with advancing levels of automation has the potential to greatly reduce human error and lessen bias and the negative impacts of emotion on human decision making processes.[6] For example, the University of California San Francisco's Medical Center uses an algorithmically operated robot to run a fully automated hospital pharmacy.[7] Forensic accounting and other financial analysis techniques are also fully operational in assisting with the detection of business manipulation of disclosed information,[8] along with protection from credit card fraud and identity theft.[9]

However, this new Big Data frontier is potentially rife with risk. For example, while Facebook allows for distant friends to stay virtually close, many entities are now using Facebook postings and communications as a means to gather information. Consider U.K.-based car insurer Admiral,

---

[3] *See, generally*, Rachael King, *How Dell Predicts Which Customers Are Most Likely to Buy*, WALL ST. J.: CIO J. (Dec. 5, 2012), https://blogs.wsj.com/cio/2012/12/05/how-dell-predicts-which-customers-are-most-likely-to-buy/ [https://perma.cc/X3PH-8GAE];
Gagan Mehra, *Predictive Analytics Is Changing eCommerce & Conversion Rate Optimization Business to Community*, BUSINESS 2 COMMUNITY, (July 27, 2014), https://www.business2community.com/big-data/predictive-analytics-changing-ecommerce-conversion-rate-optimization-0954947#h0TaBjvdfordM70D.97 (describing the process of using predictive analytics in business) [https://perma.cc/GQ5A-TBA4].

[4] "The global predictive analytics market, valued at USD 2.08 billion in 2012, is expected to see strong growth at 17.8% CAGR during 2013 to 2019." *Predictive Analytics Market - Global Industry Analysis, Size, Share, Growth, Trends, and Forecast to 2019*, (Aug. 4, 2014).PR NEWSWIRE, (April. 17, 2014), https://www.prnewswire.com/news-releases/predictive-analytics-market---global-industry-analysis-size-share-growth-trends-and-forecast-to-2019-255600791.html.

[5] In fact, data from the JiWire Mobile Audience Insights Report Q4 2011 indicates that 80% of mobile consumers prefer ads that are locally relevant to them, and three-quarters of consumers have taken action in response to a location-specific message. *See* JIWIRE, Marketing Charts Staff, 1 in 5 Mobile Users Recently Scanned QR Code, MARKETING CHARTS (Feb 2012), http://www.marketingcharts.com/wp/online/1-in-5-mobile-users-recently-scanned-qr-code-21145/ [https://perma.cc/X3PH-8GAE].

[6] *See* Martin Eiermann, *Algorithms and the Future of Work*, THE EUROPEAN (June 26, 2013), http://www.theeuropean-magazine.com/christoper-steiner/7226-algorithms-and-the-future-of-work (interview of Christopher Steiner) [https://perma.cc/5RL4-ZPTZ].

[7] *See* Karin Rush-Monroe, *New UCSF Robotic Pharmacy Aims to Improve Patient Safety*, UCSF (Mar. 7, 2011), https://www.ucsf.edu/news/2011/03/9510/new-ucsf-robotic-pharmacy-aims-improve-patient-safety [https://perma.cc/7R2Q-2LPG].

[8] *See* Messod D. Beneish, *Predicting Firms that Manipulate Disclosed Earnings*, ON ANALYTICS 6, 6-7, (Spring 2014), https://kelley.iu.edu/include/flipbook/2014springOnAnalytics/#/6/.

[9] For example, Mastercard uses an industry leader to note transactions that seem to be out of line with an individual's shopping patterns. *See* Mastercard Services, MASTERCARD, available at https://www.mastercard.us/en-us/consumers/payment-technologies/id-theft-protection.html [https://perma.cc/8ZFY-LHNP].

which attempted to launch "Firstcarquote" as a means to analyze the Facebook accounts of first-time car buyers to see whether their personalities suggest they will be safe drivers[10] and, more significantly, to adjust their rates accordingly. Fortunately, Facebook stepped in and stopped the practice. Another example is the creation of a psychological analytics model described by Michael Kranish in the *Washington Post* in which he reports on President Donald Trump's plan to build a "psychographic" profile of every voter to enable targeted campaigning.[11] Finally, consider the Facebook controversy surrounding the use of targeted advertising that seemingly allowed for marketers to exclude users by "ethnic affinity."[12]

The Artificial Intelligence (AI) community is beginning to take notice of the issues surrounding the widespread use of these technologies, especially as it relates to bias. According to James Zou, Assistant Professor for Biomedical Data Science at Stanford University, "machine systems are learning human biases."[13] For example, Princeton University researchers conducted a word association task with a popular unsupervised AI algorithm that uses online text to understand human language.[14] After training the AI through basic word associations taught by the researchers, the algorithm was then asked to create its own associations. During this unsupervised phase the AI linked white-sounding names with "pleasant" and black-sounding names with "unpleasant."[15] And these are just a few of the more recent experiences of difficulties and warnings emerging from the development and deployment of AI into our daily lives.

Perhaps the most significant stakes lie in how Big Data and AI become implicit in regulating access to basic civil and human rights. As a 2014 White

---

[10] *See* Joseph Curtis, *Facebook Blocks Car Insurer Admiral's 'Intrusive' Plan To Check Your Social Media Posts Before It Sets Premiums*, DAILY MAIL, (Nov. 2, 2016), http://www.dailymail.co.uk/news/article-3895740/Be-careful-post-Car-insurer-start-checking-Facebook-account-sets-premiums.html [https://perma.cc/28CV-7UF9].

[11] *See* Michael Kranish, *Trump's Plan For A Comeback Includes Building A 'Psychographic' Profile of Every Voter,* Washington Post (Oct. 27, 2016), https://www.washingtonpost.com/politics/trumps-plan-for-a-comeback-includes-building-a-psychographic-profile-of-every-voter/2016/10/27/9064a706-9611-11e6-9b7c-57290af48a49_story.html?utm_term=.33169cf66421 [https://perma.cc/DB7L-QPWF].

[12] *See* Stacy Liberatore, *Facebook Race Row Over Ad System That Allows Advertisers to Exclude Users by 'Ethnic Affinity'*, Daily Mail (Oct. 28, 2016), http://www.dailymail.co.uk/sciencetech/article-3883624/Facebook-race-row-ad-allows-advertisers-choose-exclude-user-ethnic-affinity.html [https://perma.cc/8HKR-PR8T].

[13] James Zou, *Are We Making AIs Racist And Sexist? Researchers Warn Machines Are Learning To Have Human Biases*, Daily Mail, (Sept. 26, 2016), http://www.dailymail.co.uk/sciencetech/article-3808834/Are-making-AIs-racist-sexist-Researchers-warn-machines-learning-human-biases.html [https://perma.cc/JR77-G4BX].

[14] *Id.*

[15] *Id.*

House Report entitled *Big Data: Seizing Opportunities, Preserving Values* notes:

> It is one thing for Big Data to segment consumers for marketing purposes, thereby providing more tailored opportunities to purchase goods and services. It is another, arguably far more serious, matter if this information comes to figure in decisions about a consumer's eligibility for—or the conditions for the provision of—employment, housing, health care, credit, or education.[16]

This Report highlights five areas of discriminatory impacts, each illustrated by well-known stories of information gathering that resulted in negative outcomes for individuals. For example, Facebook information gathering as a pre-employment screening tool appears in numerous news stories, so much so that several state legislatures have sought to limit mandatory disclosure of social website passwords.[17] Yet, while the White House Report makes the distinction between tailoring services and practicing unethical discrimination, what it leaves unclear is just how subtle, yet powerful, the unintended harms of Big Data use can be. Understanding the risks inherent in even well-intentioned uses of Big Data and machine learning, and thus crafting effective legal and policy responses, involves managers and policymakers having a critical understanding of how machine learning actually works.

In 2011 Professors Bostrom and Yudkowsky in their book *Cambridge Handbook of Artificial Intelligence* asked readers to:

> Imagine, in the near future, a bank using a machine learning algorithm to recommend mortgage applications for approval. A rejected applicant brings a lawsuit against the bank, alleging that the algorithm is discriminating racially

[16] WHITE HOUSE, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (May 1, 2014) at 57.

[17] As of May 30, 2014, legislation has been introduced or is pending in at least 28 states, and enacted in Louisiana, Maine (authorizes study), Oklahoma, Tennessee and Wisconsin. *See Employer Access to Social Media Usernames and Passwords*, NATIONAL CONFERENCE OF STATE LEGISLATURES (Nov. 2016), http://www.ncsl.org/research/telecommunications-and-information-technology/employer-access-to-social-media-passwords.aspx (last visited Feb. 3, 2017) [https://perma.cc/KLB7-WGRK]. *See generally* Ariana R. Levinson, *Social Media, Privacy, and the Employment Relationship: The American Experience*, 2 SPANISH LAB. L. & EMP. REL. J. 15, (2013) (discussing the current legislation movement). In fact, Facebook has asked for the practice to be stopped; Doug Gross, *Facebook Speaks Out Against Employers Asking For Passwords*, CNN (Mar. 23, 2012). http://www.cnn.com/2012/03/23/tech/social-media/facebook-employers/index.html. [https://perma.cc/T3JQ-TNQX]. Federal legislation has, however, stalled. *See* Sara Gates, *CISPA Amendment Banning Employers from Asking for Facebook Passwords Blocked*, HUFFINGTON POST (Apr. 23, 2013), https://www.huffingtonpost.com/2013/04/21/cispa-amendment-facebook-passwords-blocked_n_3128507.html. [https://perma.cc/MV48-LQ8R]. *See also Forty-five Percent of Employers Use Social Networking Sites to Research Job Candidates,* CAREER BUILDER (Aug. 19, 2009), http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?ed=12%2F31%2F2009&id=pr519&sd=8%2F19%2F2009 [https://perma.cc/PYT3-6YZ3].

> against mortgage applicants. The bank replies that this is impossible, since the algorithm is deliberately blinded to the race of the applicants.[18]

They go on to add: "Statistics show that the bank's approval rate for black applicants has been steadily dropping. Submitting ten apparently equally qualified genuine applicants (as determined by a separate panel of human judges) shows that the algorithm accepts white applicants and rejects black applicants."[19]

They ask: "What could possibly be happening?"[20] This scenario goes straight to the key ethical impacts outlined by the White House: discrimination in credit access is a fundamental pillar of institutional racism. Professors Bostrom and Yudkowsky use this example to highlight an often misunderstood aspect of machine learning and artificial intelligence algorithms; addressing biased outcomes is not as simple as blaming the programmer.

This article seeks to enter into the debate surrounding the need for data regulation, however, it advances the argument that any regulation of data needs to begin to encompass considerations of the use of the data and the corresponding impacts of such use. The article is structured as follows: first, it provides a brief explanation of data analytics, algorithms, and machine learning; second, the article explores some of the common mistakes associated with data modeling within algorithmic processes; third, the study explores the impact of the use of data, specifically data that is used to create a digital personhood, to inform algorithms that perform basic services; fourth and finally, the article seeks to define an ethical decision-making model and regulatory structure for data, leveraging the literature on polycentric governance and focusing on the impact of data usage upon both the individual and society.

I. ALGORITHMS, ANALYTICS, AND MACHINE LEARNING

This section provides a brief overview of algorithms and machine learning in particular to provide a foundation for discussion. At its most basic, it is important to note the distinctions between Big Data, data mining, and machine learning. Big Data is a fairly nebulous term for enormously large data sets that have only recently become possible to accumulate; by extension, it is often used colloquially to refer to the computing processes applied to those data sets and the insights thereby derived. But such

[18] Nick Bostrom & Eliezer Yudkowsky, *The Ethics of Artificial Intelligence*, CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE 516, 516 (eds. William Ramsey & Keith Frankish eds. 2014).

[19] *Id.*

[20] *Id.*

processes are not identical. Data mining, specifically, discovers previously unknown patterns and knowledge within Big Data sets. Machine learning, by contrast, is used to analyze patterns and then apply results to decision making and actions.[21] All computing processes, from mining to learning, in some sense rely on algorithms, but their scope, complexity, and conceptual accessibility varies widely.[22]

Many of the algorithms that have shaped our world so far, such as Google's PageRank and Amazon's recommendations feature, perform functions that could have previously been done with paper and pencil (and a lot of hard work). Their key contribution is that they help to order and arrange vast volumes of data at a scale and speed impossible for human beings, making increasingly large data sets "legible," and making the end user's interactive experience seamless and non-intrusive. When it comes to machine learning, however, there is a truly new element; machine learning can not only produce outputs that would not have been possible with earlier technologies, but can also be incorporated into human institutional processes in such a way that key decisions are in some sense made in the "black box" of the algorithm itself. Moreover, contrary to the common impression that if an algorithm gives a bad answer, it must be that the programmer set it up to return bad answers, it is not always so easy to deconstruct how specific outputs are produced—hence the black box. This is why any discussion of law and ethics involving machine learning must first attempt to understand the basic technical premises, and equally important, map out how machine learning and other non-algorithmic institutional actors interact to produce legally and ethically relevant decisions.

In general, the SAS Institute notes the two most widely adopted methods of machine learning are *supervised* and *unsupervised* learning:

> Supervised learning algorithms are trained using labeled examples, such as an input where the desired output is known. For example, a piece of equipment could have data points labeled either "F" (failed) or "R" (runs). The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly. Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.[23]

---

[21] Machine Learning, What it is and Why it Matters, SAS INSTITUTE, http://www.sas.com/en_us/insights/analytics/machine-learning.html (last visited Feb. 3, 2017) [https://perma.cc/P7HG-K4PY].

[22] *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 633-34 (2017).

[23] SAS INSTITUTE, *supra* note 21

Supervised learning accounts for approximately seventy percent of machine learning applications.[24] "It is commonly used in applications where historical data predicts likely future events.[25] For example, it can anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim."[26] In contrast: "Unsupervised learning is used against data that has no historical labels. The system is not told the 'right answer.' The algorithm must figure out what is being shown. The goal is to explore the data and find some structure within."[27]

The SAS Institute suggests that "[u]nsupervised learning works well on transactional data.[28] For example, it can identify segments of customers with similar attributes who can then be treated similarly in marketing campaigns.[29] Or it can find the main attributes that separate customer segments from each other."[30]

Both forms of machine learning have their strengths. Both are vulnerable to the oft-repeated GIGO principle: "garbage in, garbage out." If your data collection is limited or conceptually flawed, the conclusions the machine will spit out will be limited and flawed.[31] In the next section, we will explore some of the more subtle but potent possible problematics in machine-learning-assisted decision-making.

II. INVISIBLE BIAS AND THE MECHANISMS OF REPRODUCTION

The widespread use of algorithms has created a vast landscape of stored and networked data, and because the value of any one piece of data is not easily delimited, but rather includes the potential to yield further data through algorithmic learning processes, data on all of us is also stored and mined in ways that are frequently obscure.[32] The hidden nature of this algorithmic data

---

[24] *Id.*
[25] *Id.*
[26] *Id.*
[27] *Id.*
[28] *Id.*
[29] *Id.*
[30] *Id.*
[31] Originally coined by IBM programmer George Fuechsel, Garbage In Garbage Out (GIGO) is popular computing slang for "if you input the wrong data, the results will also be wrong." *See* What is Garbage in Garbage Out?, WISEGEEK, http://www.wisegeek.org/what-is-garbage-in-garbage-out.htm (last visited Nov. 12, 2017) [https://perma.cc/TF5R-6AH4].
[32] *See* Dan Breznitz, Seymour Goodman, & Michael Murphree, *Ubiquitous Data Collection: Rethinking Privacy Debates*, COMPUTER, at 100 (June 2011). This perpetual storage of data also increases cyber insecurity. *See* Nate Lord, *Data Security Experts Reveal the Biggest Mistakes Companies Make with Data & Information Security*, DIGITAL GUARDIAN (Oct. 12, 2016), https://digitalguardian.com/blog/data-security-experts-reveal-biggest-mistakes-companies-make-data-information-security [https://perma.cc/AYQ8-PAF8].

environment is potentially troubling and not only in terms of transparency. These are data sets in which many discrete pieces of information, together with pattern analyses made possible by algorithms, can be pieced together to complete a very accurate "virtual person." According to noted authority Eric Siegel: "[I]t is the generation of new data that can lead to the indirect discovery of unvolunteered truths about people."[33] Yet a third troubling feature of this environment is that societal assumptions based on automation bias can transform the reasonably good predictions of machine learning into accepted "truths."[34] And the manner in which these truths are created—about each of us—at the intersection of Big Data and the human social systems that utilize it, must draw the attention of ethicists and academics, lest automation bias lead us to overlook the many ways misuse, both opportunistic and naïve, can invade our daily lives.

This section explores the manner in which some of the most common pernicious impacts of invisible biases and misperceptions become engrained in algorithmic processes, with potentially devastating consequences. First, we discuss the effect of the Big Data black box on social science, which is leading many to question our ability to explore and discover potential bias in a scientific manner. Next, we explore the effects of unrecognized recursion in Big Data-driven decision-making processes, an issue we frame in terms of barriers to overcoming problematic "common sense." Finally, we look at how the existence of black box algorithms can lead to issues with "ethical agency" causing problems related to accountability and responsibility for decision making.

A. The Misuse of Statistics and Social Science

The scientific method has historically demanded adherence to a fundamental set of standardized techniques for building scientific knowledge. The use of the scientific method is the hallmark of social science research, which provides crucial insights for policymakers, educators, and law enforcement, among other groups. While certain details of an experiment may be cloaked (for example, by de-identifying human subjects), to be considered scientifically valid an experiment must be capable of replication.[35] The hidden, often proprietary nature of data sets and the algorithms that work them has led to the growth of studies that are presented

[33] Eric Siegel, PREDICTIVE ANALYTICS: THE POWER TO PREDICT WHO WILL CLICK BUY, LIE OR DIE, 38, Wiley (2013).

[34] Automation bias is an assumption that a "machine driven, software enabled system is going to offer better results than human judgment." Frank Pasquale, THE BLACK BOX SOCIETY, 107 (2015).

[35] *See* William Rand & Uri Wilensky, *Verification and Validation through Replication: A Case Study Using Axelrod and Hammond's Ethnocentrism Mode*l, Proc. NAACSOS (2006) http://ccl.northwestern.edu/papers/naacsos2006.pdf (last visited Feb. 3, 2017).

with the authority of social science, but without reporting the detailed methods necessary for scientific replication or robust critical examination. In the world of Big Data, "social science" is frequently carried out without the transparency that occurs when replication is the expected standard. As noted author, Cathy O'Neil highlights in her book, *Weapons of Math Destruction*: "Ill conceived mathematical models now micromanage the economy, from advertising to prison. . . . They're opaque, unquestioned, and unaccountable, and they operate at the scale to sort, target, or optimize millions of people."[36]

Despite the use of pseudo-science, we allow it to guide decision-making in some of the most fundamental aspects of our human existence.[37] The analytics industry[38] frequently implies that true social science, based on the scientific method and statistical analysis, is being used in the interpretations of data sets and outcomes. Proprietary data sets and black-box algorithms, however, mean that in many cases no one is able to independently verify that basic statistical principles are being followed in predictive design,[39] making it difficult to distinguish reliable findings from pseudo-science. As highlighted by Professor Shane Jensen, "I've read about many studies in medicine, economics and social science that could benefit from more discussion with statisticians about the analysis of collected data and the collection of the data itself."[40]

This situation has exacerbated the comprehension gap between experts and those using their findings, with the prestige of technology adding an aura of objectivity to analyses of widely varying quality. It is true that machine learning is a powerful tool; in particular, it is excellent at finding correlations. The dilemma, as noted by Eric Siegel writing in *Predictive Analytics*, "is, as it is often said, correlation does not imply causation. The discovery of a predictive relationship between A and B does not mean that one causes the

[36] Cathy O'Neil, Weapons of Math Destruction, 12, Crown, 2016.

[37] Consider online dating site, eHarmony. *See* Robert L. Mitchell, *Online Dating*: Analyzing the Algorithms of Attraction, PC World, (Feb. 19, 2009) https://www.pcworld.com/article/159884/online_dating_under_hood.html [https://perma.cc/PL8H-YPYS].

[38] Analytics is the discovery, interpretation, and communication of meaningful patterns in data. It is often within the area of interpretation that the line begins to blur between social science and reasoned guesses.

[39] Consider the widely criticized case of the Facebook emotion contagion study, in which the absence of informed consent (and the ability to opt out or quit the study) was unavailable to unknowing participants. *See* Inder M. Verma, *Editorial Expression of Concern and Correction*, 111 PROC. NAT'L ACAD. SCI. U.S., 10778 (2014) *responding to* Adam D. I. Kramer, Jamie E. Guillory, & Jeffrey T. Hancock, 111 PROC. NAT'L ACAD. SCI. U.S. (2014).

[40] Ric Bradlow, Shane Jensen, Justin Wolfers & Adi Wyner, *Report Backing Clemens Chooses Its Facts Carefully*, N.Y. Times (Feb. 10, 2008), http://www.nytimes.com/2008/02/10/sports/baseball/10score.html [https://perma.cc/NWW7-XN82].

other, not even indirectly. No way, no how."[41] Making this point in a humorous manner, Professor David Leinweber[42] found questionable correlations amongst financial indicators,[43] and used this contrived analysis to form an elaborate publicity stunt in which he stated the "best predictor of the S&P 500's performance" is the production of butter in Bangladesh.[44] The global financial crisis put predictive analytics to the test; a test that, some argue, was an abject failure due to incorrect assumptions that existed within the system.[45] As *New York Times* author Saul Hansell explains: "Financial firms chose to program their risk-management systems with overly optimistic assumptions. . . . Wall Street executives had lots of incentives to make sure their risk systems didn't see much risk at all."[46] Yet, of the many financial analysts who use simulated investment results as a means to test investment performance, it seems that few fully appreciate how the logic underlying these simulations has frequently been problematic.

This absence of the adherence to the scientific method as a cornerstone within digital social science potentially removes legally and professionally created procedures and protections that have long protected the ethical integrity of the field and been considered hallmarks of the profession. For example, consider the Facebook research scandal in which Facebook manipulated nearly 700,000 users' news feeds to see whether it would affect their emotions.[47] Critics and ethicists argued that the absence of informed consent—considered an expectation within the social science research community—was ignored within this experimental design.[48] According to one of the researchers, Adam Kramer, this type of social science should be

---

[41] Siegel, *supra* note 33, at 88.

[42] *See* David Leinweber, *Stupid Data Miner Tricks*, NERDS ON WALL STREET, (2009), http://nerdsonwallstreet.typepad.com/my_weblog/files/dataminejune_2000.pdf

[43] *See* Siegel, *supra* note 33, at 120.

[44] Selena Maranjian, *Butter in Bangladesh Predicts the Stock Market*, THE MOTLEY FOOL (Sept. 20, 2007), https://www.fool.com/investing/general/2007/09/20/butter-in-bangladesh-predicts-the-stock-market.aspx [https://perma.cc/NTU7-MMAK].

[45] *See* John D. Freeman, *Behind the Smoke and Mirrors: Gauging the Integrity of Investment Simulations*, FIN. ANALYSTS J., Nov.-Dec. 1992 at 26.

[46] Saul Hansell, *How Wall Street Lied to Its Computers*, N.Y. TIMES, (Sept. 18, 2008), https://bits.blogs.nytimes.com/2008/09/18/how-wall-streets-quants-lied-to-their-computers/?dbk [https://perma.cc/W9FV-NVSG].

[47] *See e.g.*, Kashmir Hill, *Facebook Manipulated 689,003 Users' Emotions For Science,* FORBES (June 28, 2014), https://www.forbes.com/sites/kashmirhill/2014/06/28/facebook-manipulated-689003-users-emotions-for-science/#301ce19d197c [https://perma.cc/78BE-H6BY]; Robert Booth, *Facebook Reveals News Feed Experiment To Control Emotions*, THE GUARDIAN, (June 30, 2014), https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds [https://perma.cc/4FHC-2QME].

[48] Charles Arthur, *Facebook Emotion Study Breached Ethical Guidelines, Researchers Say,* THE GUARDIAN (June 30, 2014), https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say [https://perma.cc/W3TS-DRZD].

allowed "because we care about the emotional impact of Facebook and the people that use our product."[49] The statement demonstrates a common concern of "online social science," which is that there are often an inadequate number of academically trained traditional social scientists involved. More collaboration between academics and industry would help promote informed consent.[50] However, as noted by Wharton researchers Ric Bradlow, Shane Jensen, Justin Wolfers, and Adi Wyner, when it comes to "statisticians-for-hire," there is a tendency to choose comparison groups that support their clients.[51]

When unevaluated, questionably unethical (or at least lacking true consent) pseudo social science is allowed to be relied upon in the same manner as true social science, regulators must begin to consider the impact of outcomes on society as a whole.

B. Engraining of "Common Sense"

Common sense and common knowledge often run hand in hand and neither has a universal definition that transcends generations or ethical debate. Despite the lack of clear definition, one thing is certain—regardless of the different shades of meaning, the use of the term "common sense" implies education and wisdom. Yet, within that shading is an understanding that common sense is shared by—or "common to" —nearly all people without any need for debate.[52] One then must wonder if we should be content with common sense being captured and memorialized within an algorithm. Consider gendered advertising. The color pink has been associated with girls for barely a hundred years,[53] yet at this point in time the association is so powerful that it has entered common sense that a pink product is a product for women, something many consumers in the West know and reflect in their behavior, regardless of how they actually feel.[54] There are many criticisms that can be made of such crude readings of gender, including that such conventions may serve to limit market creativity by replacing actual

---

[49] *See id.*

[50] *See* General Requirements for Informed Consent, 45 C.F.R. § 46.116 (2009).

[51] Bradlow et al., *supra* note 40.

[52] *See* FRITS VAN HOLTHOORN & DAVID R. OLSON, COMMON SENSE: THE FOUNDATIONS FOR SOCIAL SCIENCE (1987) ("common sense consists of knowledge, judgment, and taste which is more or less universal and which is held more or less without reflection or argument").

[53] *See* Claudia Hammond, *The Pink versus Blue Gender Myth*, BBC (Nov. 14, 2014), http://www.bbc.com/future/story/20141117-the-pink-vs-blue-gender-myth [https://perma.cc/94U8-6HTH].

[54] *See* Natalie Wolchover, *Why Is Pink for Girls and Blue for Boys?,* LIVE SCIENCE (Aug. 1, 2012), https://www.livescience.com/22037-pink-girls-blue-boys.html (discussing the gender norms that guide decision making) [https://perma.cc/GG38-VEBK].

experimentation, research, and innovative development with a dependable and predictable women's "market" created more or less by fiat.[55] What are we to do with this type of common sense that should no longer be viewed as common sense? One of the positive potentials of Big Data in advertising is the ability to actually get beyond these limiting and biased models to better identify and better serve diverse populations. However, unrecognized recursion poses a challenge to fully unlocking this potential.

To understand the challenge, consider the popular game, the Tower of Hanoi. This traditionally wooden puzzle consists of three rods, and a number of disks of different sizes that can slide onto any rod. The puzzle starts with the disks in a neat stack in ascending order of size on one rod, the smallest at the top. The objective of the puzzle is to move the entire stack to another rod while obeying a set of simple rules while learning basic problem-solving skills. As many players figure out, there is a pattern that, if repeated, will result in the objective being achieved. As such, the Tower of Hanoi may be considered a popular way to teach recursive algorithms to a beginning programming student. Recursion in computer science is a method in which the solution to a problem depends on solutions to smaller instances of the same problem.[56] This method makes larger, often multistep problems more manageable and capable of compounding solutions. However, by the same token, when a mistake is included in a recursive process, its repetition leads to compounded problems.

We argue that when machine learning is included in a decision-making process that also includes institutional and human actors, we can analyze that entire process as a learning system: the feedback of information from data through algorithmic processing into information outputs interpreted by human/institutional actors, into social impacts, and back into data again. Zooming out from the machine-learning component to view the overall learning system, we can analyze ways it is susceptible to both the benefits and risks of unrecognized recursion.

Perhaps no case illustrates these dangers more than predictive policing. High-tech policing is no longer the storyline in a movie; it is a reality in many communities such as New York, Houston, Seattle, and Fresno.[57] And the

---

[55] *See* CLAIRE CAIN MILLER, BOYS AND GIRLS, CONSTRAINED BY TOYS AND COSTUMES, N.Y. TIMES (Oct. 30, 2015), https://www.nytimes.com/2015/10/31/UPSHOT/BOYS-AND-GIRLS-CONSTRAINED-BY-TOYS-AND-COSTUMES.HTML (discussing the gender division in childhood). [https://perma.cc/A3XT-AN38].

[56] *See* Behrouz A. Forouzan, FOUNDATIONS OF COMPUTER SCIENCE, Cengage Learning EMEA, (3rd Rev. ed.) (December 5, 2013).

[57] Justin Jouvenal, *The New Way Police Are Surveilling You: Calculating Your Threat 'Score',* WASH. POST (Jan. 10, 2016), https://www.washingtonpost.com/local/public-safety/the-new-way-police-

level of data gathering and analytics, even prediction, may surprise some. For example, in the Real Time Crime Center in Fresno the software application known as Beware, scours "billions of data points, including arrest reports, property records, commercial databases, deep Web searches" as well as a citizen's "social-media postings."[58] These data points are then used to calculate a threat level that is fed back to the officers to be used in their decision-making process, often in real time volatile situation.[59] Many communities have particular aspects of high-tech policing. While the most common forms of surveillance are cameras and automated license plate readers, the use of handheld biometric scanners, social media monitoring software, devices that collect cellphone data, and drones are all being used as surveillance tools.[60] And all of this data is fed into larger databases that expand the information available to governments. For example, the FBI's Next Generation Identification (NGI) project collects fingerprints, iris scans, data from facial recognition software, and other sources that aid local departments in identifying suspects.[61]

There certainly is room for improved policing, but these scenarios are cause for concern. The Ford Foundation's Michael Brennan lays out the problem:

> I recently attended a meeting about some preliminary research on 'predictive policing,' which uses these machine learning algorithms to allocate police resources to likely crime hotspots. The researchers at the Human Rights Data Analysis Group discussed how these systems are often deployed in response to complaints about racial bias in policing, but the data used to train the algorithm comes from the outcomes of the biased police activity.[62]

As we've seen with earlier examples, here, the outcomes of machine learning may be illuminating, but it is crucial to keep in mind that *the machine got its information from us*. Therefore ethically relevant data corruption originates long before the machine learning program is written. The data fed in, even if uncorrupted from a data science point of view, is itself a record of human bias. For vendors and policymakers, Big Data is an exercise in finding legitimate insight in the echo chamber. For individuals,

---

are-surveilling-you-calculating-your-threat-score/2016/01/10/e42bccac-8e15-11e5-baf4-bdf37355da0c_story.html?utm_term=.eb77e8224833 [http://perma.cc/SW46-TDBT].

[58] *Id.*

[59] *See id.*

[60] *See id.*

[61] *See id. See also* Next Generation Identification (NGI), FBI, https://www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi (last visited February 26, 2016) [http://perma.cc/SQ9B-298B].

[62] Cory Doctorow, *Racist Algorithms: How Big Data Makes Bias Seem Objective*, BOINGBOING, (Dec.2, 2015), https://boingboing.net/2015/12/02/racist-algorithms-how-big-dat.html [http://perma.cc/S9UA-U9FF].

both as consumers and citizens, the line between accessing information and submitting to conditioning becomes ever thinner. If you have ever used someone else's computer without ad blocking, you have seen how, despite increasing ad sophistication, there remains an intense blunt force effect from bots getting one or two simple pieces of data, and shaping the world that you experience.

Brennan goes on to unpack the implications of predictive policing in chilling terms:

> If the police are stop-and-frisking brown people, then all the weapons and drugs they find will come from brown people. Feed that to an algorithm and ask it where the police should concentrate their energies, and it will dispatch those cops to the same neighborhoods where they've always focused their energy, but this time with a computer-generated racist facewash that lets them argue that they're free from bias.[63]

Thus, the second important point is that *we are now learning from the machines.* This is where problems like the already discussed issue of imagined objectivity, and our next topic of concern, elided agency, come into play.

C. Elided Agency

One of the common issues that arise when discussing invisible biases that become engrained in algorithmic processes arises from the unwillingness of individuals or institutions to acknowledge the existence of bias due to the 'it was the algorithm' justification. This issue often arises because our notions of accountability and responsibility are susceptible to blur when decisions come to be made by algorithms. The case that has brought this issue to the forefront of public attention is the much-discussed self-driving vehicle.[64] If a self-driving vehicle finds itself in a situation where there is no course of action that will not cause harm to human life—for example, it can hit a child on the road or swerve off the edge of the cliff, risking the driver's life – how do we apply concepts of manslaughter and murder? Various discussions have considered how to apportion liability among the driver, the programmers (a large group), the manufacturer, and the software proprietor.[65] One aspect of this problem that these discussions

---

[63] *See id.*

[64] For an excellent example of the dilemma and a discussion about the issue, see the Massachusetts Institute of Technology Lab at MIT Media Lab at http://moralmachine.mit.edu/

[65] *See e.g.,* Alexis C. Madrigal, *If a Self-Driving Car Gets in an Accident, Who—or What—Is Liable?*, THE ATLANTIC (Aug. 13, 2014), http://www.theatlantic.com/technology/archive/2014/08/if-a-self-driving-car-gets-in-an-accidentwho-is-legally-liable/375569 [https://perma.cc/73KR-ESKH]; *Who is responsible for a driverless car accident?*, BBC NEWS (Oct. 8, 2105), http://www.bbc.com/

do not always cover is the issue of ability and responsibility: algorithms have capacities that differ from humans. A human driver in this situation does not actually have time to make a decision, but reacts on reflex, and a human driver may be guilty of taking various risks that contribute to an accident, such as driving while impaired.[66] There are many cases in which a self-driving car has the capacity to perceive and enact a course of action that a human brain would not be able to compute; on the flip side, they can be highly predictable in ways humans cannot—it is safe to assume that neither speeding nor texting contributed to an accident by a self-driving car, for example.[67] In other words, the considerations that bear on the difference between murder and manslaughter may not be the same for a learning machine as for a human. Of course, cars do not have agency in the sense of consciousness. Nonetheless, to the extent that they are capable of making determinations, including novel determinations that no programmer foresaw or intended, we may need to adjust our notions of ethical responsibility.

The ethical and liability questions of self-driving vehicles are already widely recognized. What deserves more attention are processes in which machine decision-makers are either replacing human decision-makers, or introducing entirely new decision points into a human process, without it being realized by those affected. In such cases, there is a danger of elided agency—outcomes that once were someone's responsibility can appear inevitable or fade into background conditions when in fact, there is a legally and/or ethically relevant juncture, with no ethically relevant human actor in it—because a machine was the relevant decision-maker.

For example, although the traditional concept of censorship implies conscious, ideological agency, unconscious automated filters can produce results that look very much like censorship. Many commentators argue that algorithm censorship is growing at an alarming rate—and yet many citizens remain unaware of its existence. Consider the simple, yet common, example

news/technology-34475031 [https://perma.cc/4NGQ-2FYX]; Stephanie Hsu, *Self-Driving Cars and Liability Implications,* FORDHAM INTELL. PROP., MEDIA & ENT. L. J. BLOG (Sept. 16, 2016) http://www.fordhamiplj.org/2016/09/16/self-driving-cars-liability-implications/ [https://perma.cc/GVT8-64SM].

[66] Patrick Lin, *The Ethics Of Autonomous Cars*, The Atlantic (Oct. 8, 2013), http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360 (discussing the self-driving car dilemma and if programming should actually 'act like a human' with all of his deficits). *See also* Joe Myers, *How Will Self-Driving Cars Make Life or Death Decisions?*, World Economic Forum (Aug. 15, 2016), https://www.weforum.org/agenda/2016/08/the-ethics-of-self-driving-cars-what-would-you-do/ [https://perma.cc/8BHE-J8MP]; Karen Kaplan, *Ethical Dilemma on Four Wheels: How to Decide When Your Self-Driving Car Should Kill You*, L.A. Times (June 23, 2016, 11:05 AM), http://www.latimes.com/science/sciencenow/la-sci-sn-autonomous-cars-ethics-20160623-snap-story.html [https://perma.cc/29JM-BCPQ].

[67] Lin, *supra note* 66.

of the blocking of the livestream of the 2012 Democratic National Convention.[68] On September 4, 2012, the DNC posted on YouTube, an official streaming partner, several videos of speeches and other highlights of the DNC's evening events, all of which were featured prominently on BarackObama.com and the YouTube channel DemConvention2012. Some portion of the DNC videos triggered the YouTube digital fingerprinting system and, as a result, YouTube put a copyright blocking message on the livestream video.[69] Even more concerning in this instance are the multiple parties that claimed copyright infringement,[70] and the ability of a single (alleged) copyright holder to shut down a live stream of a publicly important event.[71] The blocking is demonstrative of the limitations of algorithms in some key areas; as *Wired* magazine author Andy Baio explains: "The inability to understand context and parody regularly leads to 'fair use' videos getting blocked, muted or monetized."[72] Moreover, as technology expert Parker Higgins notes: "It's impossibly complicated to define in a set of 'business rules' for automated [copyright] enforcement."[73]

Although in this paper we are focusing on danger zones, it is important to note at this juncture that the use of algorithms can be highly positive. Machine learning can offer a means to uncover and overcome existing, often hidden, biases, enhance human/institutional decision-making, and reduce error in areas of human agency. For example, a recent paper from the National Bureau of Economic Research comes to the conclusion that "relying on a 'feel' for a candidate—as opposed to objective qualifications—

---

[68] *See* Ryan Singel, *YouTube Flags Democrats' Convention Video on Copyright Grounds*, WIRED (Sept. 5, 2012), http://www.wired.com/threatlevel/2012/09/youtube-flags-democrats-convention-video-on-copyright-grounds/ [https://perma.cc/7NZR-EAC7]; Geeta Dayal, *The Algorithmic Copyright Cops: Streaming Video's Robotic Overlords*, WIRED (Sept. 6, 2012), http://www.wired.com/threatlevel/2012/09/streaming-videos-robotic-overlords-algorithmic-copyright-cops [https://perma.cc/Q9JV-GHAC].

[69] *See* Singel, YouTube Flags, *supra note* 68.

[70] The message that appeared instead of the video is clear:

> This video contains content from WMG, SME, Associated Press (AP), UMG, Dow Jones, New York Times Digital, The Harry Fox Agency, Inc. (HFA), Warner Chappell, UMPG Publishing and EMI Music Publishing, one or more of whom have blocked it in your country on copyright grounds. Sorry about that.

Id.

[71] *See* Dayal, *supra* note 68.

[72] Andy Baio, *Copyright Kings Are Judge, Jury and Executioner on YouTube*, WIRED, (Feb. 29, 2012), http://www.wired.com/2012/02/opinion-baiodmcayoutube [https://perma.cc/LE8Q-RH4R].

[73] Parker Higgins, *Mars Landing Videos, and Other Casualties of the Robot Wars*, ELECTRONIC FRONTIER FOUNDATION (Aug. 8, 2012), https://www.eff.org/deeplinks/2012/08/mars-landing-videos-andother-casualties-robot-wars.

makes managers' hiring decisions worse."[74] To make this determination, the researchers introduced a series of job tests in the hiring process.[75] However, the managers were allowed to either use or override the testing program's suggestions.[76] According to the study, when managers used their discretion to override the hiring order implied by the test results, the outcomes, in terms of both tenure and productivity, were worse.[77]

In this case, because the human managers still take the final action with the advisement of the algorithm, they clearly assume responsibility, ethically and legally, for that action. However, increasingly, automated programs are removing human actors from the decision-making process entirely, raising knottier questions.

In summary, we can conclude that the use of Big Data and machine learning technologies offers amazing opportunities to supplement and surpass human decision-making, but also poses a risk of reinforcing pre-existing problems through recursive learning between humans and machines. Invoking words like "data" and "algorithm" can provide a veneer of objectivity and social science that prevents the critical discussion and iterative improvement necessary to actually use them wisely. Thus, far from being a shortcut, this technology demands that police, programmers, policymakers, and citizens bring a high level of social, historical, and statistical literacy to interpreting these data responsibly.[78]

III. SOCIETAL IMPACTS OF BIAS

It is difficult to argue against the assertion made by *Guardian* writer Leo Hickman that "algorithms rule the world."[79] Algorithms can be a force for good. They can be used to locate people in difficulty, such as those in financial distress, and can be used to improve the safety of all of us, such as no-fly lists.[80] Conversely, when used improperly, algorithms can further isolate the already marginalized from capital opportunity, or recursively enhance discrimination against historically vulnerable groups. In this

[74] Mitchell Hoffman, Lisa B. Kahn, & Danielle Li, *Discretion in Hiring*, (NBER Working Paper No. 21709, Issued in Nov. 2015), http://www.nber.org/papers/w21709

[75] *See id.*

[76] *See id.*

[77] *See id.*

[78] *See id.*

[79] Leo Hickman, *How Algorithms Rule The World*, GUARDIAN (July 1, 2013). https://www.theguardian.com/science/2013/jul/01/how-algorithms-rule-world-nsa [https://perma.cc/9EVV-MJGX].

[80] Byron Tau, *No-Fly List Is Only One of Many U.S. Watchlists*, WALL STREET J. (Dec. 8, 2015), https://www.wsj.com/articles/no-fly-list-is-only-one-of-many-u-s-watchlists-1449570602 [https://perma.cc/N3CW-PV93].

section, we lay out some of the major areas of concern for policymakers trying to guide the ways in which these tools are used.

A. Barriers to Economic Development and Basic Life Services

One of our major areas of concern with algorithmic processes are biases that reinforce existing social discrimination. For example, we know that institutional racism enacted in law and policy by local governments, big banks, and big business—not merely by individuals exercising their individual racist preferences—has already created areas of low service, limited capital, cyclical debt, and population disruption within cities.[81] We have already seen several examples of how machine learning can "pick up" racism and sexism.[82] The consequences can be dire if the machine learning is feeding into systems like housing, healthcare, market access, and transit.

Several new uses of Big Data promise to tie credit ratings—already a rigged system—into even more discriminatory barriers. According to a recent story by *Forbes*,[83] credit rating agencies may soon be harvesting data on word usage on social media to adjust credit scores. And other reports suggest that lenders are looking at social networks themselves[84]—i.e., who you know—which would be an enormous disadvantage for people from low-credit communities, even if their individual credit-worthiness is high.

In the area of economic disadvantage and new technology, we see a lot of old wolves in new sheepskins. For example, even as Uber is challenged on its status as an employer,[85] one commenter argues that status needs to be regulated in the interest of consumers, to ensure equal access to the transportation services:

> Currently, Uber and Lyft are killing cabs because [they] are cheaper and cooler. Part of the reason [they're] cheaper is because [they] aren't taxed and regulated as highly. Part of the reason you're cooler is because Lyft and Uber

---

[81] *See* Jeff Nesbit, *Institutional Racism Is Our Way of Life*, US NEWS (May 6, 2015), https://www.usnews.com/news/blogs/at-the-edge/2015/05/06/institutional-racism-is-our-way-of-life [https://perma.cc/CDQ8-N3G8].

[82] *See generally,* Latanya Arvette Sweeney, *Discrimination in Online Ad Delivery*, 56 COMM. ASS'N COMPUTING MACHINERY 5, (2013); Amit Datta, Michael Carl Tschantz, and Anupam Datta, *Automated Experiments on Ad Privacy Settings*, PROC. ON PRIVACY ENHANCING TECH. 2015, 92–112 (2015).

[83] Bill Hardekopf, *Your Social Media Posts May Soon Affect your Credit Score*, FORBES (Oct. 2, 2015), https://www.forbes.com/sites/moneybuilder/2015/10/23/your-social-media-posts-may-soon-affect-your-credit-score-2/#78d57784f0e4 [https://perma.cc/YD5Z-VZGW].

[84] Dan Tynan, *How Facebook Can Hurt Your Credit Rating*, PC WORLD, (2011), https://www.pcworld.com/article/246511/how_facebook_can_hurt_your_credit_rating.html [https://perma.cc/S2PZ-98MS].

[85] *See* Steve Hargreaves, *Uber Driver Is, In Fact, An Employee*, CNN MONEY (June 17, 2015), http://money.cnn.com/2015/06/17/technology/uber-employee-ruling/index.html [https://perma.cc/T5PK-ES67].

> drivers are generally English-speaking, friendly, helpful young men and women. I see the appeal, but since you are currently free to not serve my neighborhood, it means that the taxis that you are killing are my only option and I'm watching it wither away. . . . Why are so many otherwise thoughtful decent people swinging on the Uber and Lyft bandwagons when they are, in essence, facilitating a return to institutional racism by choosing a service that is specifically not obligated by the protections we as a society agreed many years ago were necessary in order to preserve equal access?[86]

While new business models raise the question of new regulation discussed further in Part IV, in the case of sharing economy apps that claim to be "transparent" and "empowering" it can be all too easy to overlook old problems, like redlining, in new guises.

B. Power Differentials & Economic Incentives in the Information Marketplace

As noted authority Frank Pasquale writes in his *The Black Box Society,* "Knowledge is power. To scrutinize others while avoiding scrutiny oneself is one of the most important forms of power."[87] In an economy where data is exponentially valuable, but only large institutions have the means to mine that value, individuals are increasingly incentivized to give away large amounts of their personal data.

While many such instances of this incentivization may be quite reasonable and benign, the use of leverage to obtain profitable data occurs in areas where broader considerations of justice may apply, such as basic life services. Consider John Hancock Insurance, which announced in April 2015 that it would partner with Vitality to encourage a healthier lifestyle and offer insurance discounts for those committed to a healthy lifestyle.[88] However, as noted cybersecurity scholar Evgeny Morozov noted*:* "Unlike the rich, who pay for their connectivity with their cash, the poor pay for it with their data."[89]

A story in *PC World* offers a similar example from another sector of basic services; lending.

---

[86] Xajaxsingerx, *Ridesharing and Redlining: Uber, Lyft, Race and Class,* DAILY KOS: EMINENTLY CREDULENT MUSINGS (May 27, 2014), https://www.dailykos.com/stories/2014/5/27/1302417/-Ridesharing-and-Redlining-Uber-Lyft-Race-and-Class.

[87] Frank Pasquale, THE BLACK BOX SOCIETY 3 (2015).

[88] Tara Siegel Bernard, *Giving Out Private Data for Discount in Insurance*, N.Y. TIMES, April 8, 2015, at B1.

[89] Evgeny Morozov, *Facebook Isn't A Charity: The Poor Will Pay By Surrendering Their Data*, GUARDIAN (April 25, 2015), https://www.theguardian.com/commentisfree/2015/apr/26/facebook-isnt-charity-poor-pay-by-surrending-their-data [https://perma.cc/MK39-3VRV].

> Hong Kong-based micro-lender Lenddo – which asks for your Facebook, Twitter, Gmail, Yahoo, and Windows Live logons when you sign up -- reserves the right to rat you out to all your friends. Per Lenddo's FAQ: As long as you don't fall behind on any Lenddo loan installments, you have complete control over your privacy settings and your information will only be shared with your permission. IF YOU FAIL TO REPAY, Lenddo MAINTAINS THE RIGHT TO NOTIFY YOUR FRIENDS, FAMILY, AND COMMUNITY.[90]

The market for micro-loans is focused on the economically disadvantaged. This is a case of an explicit tradeoff: less privacy for access to credit that is otherwise unavailable.

A frequent refrain in discussions of online data use and privacy is that consumers "choose" to give up their data in return for service and convenience.[91] One criticism is that many consumers are not sufficiently tech literate to realize how much they are giving away. But equally important, data is the price of admission for many basic services. While you might choose not to have a Facebook page, when everyone from employers to pizza delivery places demand your demographic data, it is an enormous logistical and psychological challenge to find a way to say no while still accessing basic services. There is limited ethnographic evidence that certain literate consumer groups do make choices about whether to use specific services based on whether they feel they are getting equivalent value for their personal data.[92] Yet at the end of the day, even the most literate and attentive consumers would have a hard time keeping up with constant unilateral end user license agreement updates on the software that powers their lives.[93]

In many ways, there is no more concerning area of knowledge and power dissymmetry than in the health care field. The urge to apply insights from Big Data and Machine Learning can complicate existing inequities in access to medical care and personal choice—a hot topic in Big Data circles.[94] Big Data *used correctly* may improve our ability to prevent epidemics and simultaneously enhance individual outcomes through personalized medicine. For example, Stanford researchers are working on the creation of

---

[90] Tynan, *supra* note 84 (emphasis in original).

[91] *See id.*

[92] *See* Lee Rainie & Maeve Duggan, *Privacy and Information Sharing*, PEW RES. CTR (Jan. 14, 2016), http://www.pewinternet.org/2016/01/14/privacy-and-information-sharing/ [https://perma.cc/L882-RUVM].

[93] *See generally* Anjanette H. Raymond, *The Consumer As Sisyphus: Should We Be Happy With 'Why Bother' Consent?*, J. OF LEGAL STUD. IN BUS., Vol. 20 (2016); Anjanette H. Raymond, *Yeah, But Did You See the Gorilla? Creating and Protecting an 'Informed' Consumer In Cross Border Online Dispute Resolution,* 19 HARV. NEGOT. L. REV., 129 (2014).

[94] *See* Bernard Marr, *How Big Data Is Changing Healthcare*, FORBES (April 21, 2015), https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#d15ecda28730 [https://perma.cc/8YHU-LH49].

an algorithm that identifies individuals with familial hypercholesterolemia,[95] which is a disorder that drastically increases the likelihood of suffering a heart-related incident, yet is often hidden.[96] Unfortunately, the cluster of behaviors and symptoms that underlie this disease tend to be difficult to identify, prior to a heart related incident.[97] Thus, the Stanford researchers hope to create an algorithm that allows for early detection.[98]

However, *who* has the responsibility, or the right, to use these data to improve health outcomes is a major question. For example, Castlight Health, a third party health care management app for employees, gathers user information from various sources, including medical information reported by individuals in the medical management system, and reports the information back to employers.[99] According to the *Wall Street Journal*: "[A]fter finding that 30% of employees who got second opinions from top-rated medical centers ended up forgoing spinal surgery, Wal-Mart tapped Castlight to identify and communicate with workers suffering from back pain."[100] Jonathan Rende, Castlight's chief research and development officer stated: "Castlight then contacted employees whose insurance and drug claims included back problems, painkillers and spinal injections with advice for physical therapists or second opinions."[101] This attempt to nudge employees,[102] the more vulnerable party in the interaction because they rely

[95] "A genetic disorder that causes high levels of LDL cholesterol, so called 'bad cholesterol,' in the blood starting in utero." Anna Maria Barry-Jester, *An Algorithm Could Know You Have A Genetic Disease Before You Do*, FIVETHIRTYEIGHT (Jan. 13, 2016), https://fivethirtyeight.com/features/an-algorithm-could-know-you-have-a-genetic-disease-before-you-do/ [https://perma.cc/9C89-SRBY]

[96] *See id.*

[97] *See id.*

[98] *See id.*

[99] CASTLIGHT HEALTH. http://www.castlighthealth.com/. Health care data mining is a growing market. *See* Adam Tanner, *This Little-Known Firm Is Getting Rich Off Your Medical Data*, FORTUNE (Feb. 9 2016), http://fortune.com/2016/02/09/ims-health-privacy-medical-data/ [https://perma.cc/SD5C-QX2F].

[100] Rachel Emma Silverman, *Bosses Tap Outside Firms to Predict Which Workers Might Get Sick*, WALL STREET J. (Feb. 17, 2016), https://www.wsj.com/articles/bosses-harness-big-data-to-predict-which-workers-might-get-sick-1455664940 [https://perma.cc/4N8H-RDV3].

[101] Under the Health Insurance Portability and Accountability Act of 1996, employers are not allowed to view their employees' health information. But that doesn't apply to third parties like Castlight. In addition, because the 'Big Data' seen by the employers only shows numbers, not individuals it is not forbidden by the 1996 act. See Valentina Zarya, Em*ployers Are Quietly Using Big Data to Track Employee Pregnancies*, FORTUNE, (Feb. 17 2016) http://fortune.com/2016/02/17/castlight-pregnancy-data/ [https://perma.cc/SD5C-QX2F].

[102] "'Nudging' involves structuring the choices that people make so as to lead them towards particular outcomes." Robert Baldwin, *Nudge: Three Degrees of Concern*, LSE Policy Briefing Paper No. 7 (2015); Timothy L. Fort, Anjanette H Raymond & Scott J. Shackelford, *The Angel on Your Shoulder: Prompting Employees to Do the Right Thing Through the Use of Wearables*, 14 NW. J. TECH. & INTELL. PROP. 139, 170 (2016).

on the employer and insurer for access to healthcare, to make medical decisions they never intended, raises questions of where the line lies between caring and interfering.

C. Influencing the Information Received

Issues of censorship—couched in intellectual property rights—also exist within the realm of search engines.[103] For example, in January 2012, the British Recorded Music Industry (BPI), Motion Pictures Association (MPA), Producers Alliance for Cinema and Television (PACT), the Premier League, and the Publishers Association proposed an "Anti-Piracy Code of Practice for Search Engines."[104] The proposed code requires search engines to demote sites in their rankings for repeatedly making available pirated content. To facilitate such a change, search engines would be expected to promote within their rankings "licensed" or "certified" websites.[105] Search engines would also agree to stop promoting pirate websites, to not place ads on those sites, and refrain from selling keyword advertising related to piracy terminology.[106] In addition to protecting intellectual property rights, potentially at the risk of censoring some kinds of information altogether, algorithms are also daily tailoring what each of us sees on the internet, to the point that users have increasingly fewer points of common reference, and some users may be effectively censored from ever accessing certain types of content at all. Few people are surprised to learn that Google alters your search results and Amazon changes the products it shows you based on products you have viewed and purchased.[107] Many consumers find this to be a benefit. Yet, they might be surprised to learn that prices can also be tailored.[108]

More troubling, the manner in which we receive news is controlled, in many ways, by an algorithm.

---

[103] *See* Anjanette Raymond, *Heavyweight Bots in the Clouds: The Wrong Incentives and Poorly Crafted Balances That Lead to the Blocking of Information Online*, 11 NW. J. TECH. & INTELL. PROP. 473, 482 (2013); Anjanette Raymond, *Intermediaries Precarious Balance within Europe: Oddly Placed Cooperative Burdens in the Online World*, 11 NW. J. TECH. & INTELL. PROP. 359, 381 (2013).

[104] Pinsent Masons LLP, *Anti-Piracy Code Of Practice For Search Engines Proposed By Rights Holder Representatives* (Jan. 27, 2012), http://www.out-law.com/en/articles/2012/january-/anti-piracy-code-of-practice-for-search-engines-proposed-by-rights-holder-representatives/ [https://perma.cc/75DV-C9LL].

[105] *See* Executive Summary, *Open Rights Group, Responsible Practices for Search Engines in Reducing Online Infringement Proposal for a Code of Practice*, *available at* http://www.openrightsgroup.org/assets/files/pdfs/proposals%20to%20search%20engines.pdf. Unsurprisingly, many of the largest search engines have taken issue with being forced to shoulder such heavy burdens instead of sharing the burden with rights holders. *See id.*

[106] *See id.*

[107] *See id.*

[108] *See id.*

> Algorithms are basically everywhere in our news environment, whether it's summarization, personalization, optimization, ranking, association, classification, aggregation, or some other algorithmic information process. Their ubiquity makes it worth reflecting on how these processes can all serve to systematically manipulate the information we consume, whether that be through embedded heuristics, the data fed into them, or the criteria used to help them make inclusion, exclusion, and emphasizing decisions.[109]

Fake newsfeed articles became a topic of discussion in the 2016 presidential election.[110] As President Obama argued: "People, if they just repeat attacks enough, and outright lies over and over again, as long as it's on Facebook and people can see it, as long as it's on social media, people start believing it. And it creates this dust cloud of nonsense."[111] It is not as if the generators of fake-news stories had social science in mind; instead they elected to use social science to an unethical end, justified by economic incentives. As *The Washington Post* reported, Paul Horner makes $10,000 a month off of ads on his pro-Trump fake-news stories. As he explained: "My sites were picked up by Trump supporters all the time . . . I think Trump is in the White House because of me. His followers don't fact-check anything—they'll post everything, believe anything."[112]

Unfortunately, Mr. Horner's surprise is unsurprising to many; social scientists have been examining the role of social media role in creating bubbles of influence for some time. In fact, previous research even explored the potential impact of social media on political developments. In 2014, Cornell Computational Communication Lab researchers noted "that media events not only generate large volumes of tweets, but they are also associated with (1) substantial declines in interpersonal communication, (2) more highly concentrated attention by replying to and retweeting particular users, and (3) elite users predominantly benefiting from this attention."[113] Using "eight major events during the 2012 U.S. presidential election, they examined "how patterns of social media use change during these media

---

[109] Nick Diakopoulos, *Nick Diakopoulos: Understanding Bias In Computational News Media,* NEIMAN LAB (Dec. 10, 2012), http://www.niemanlab.org/2012/12/nick-diakopoulos-understanding-bias-in-computational-news-media/ [https://perma.cc/U2BD-A43D].

[110] *See* Maya Kosoff, *Obama Slams Facebook for "Outright Lies" Online,* VANITY FAIR, (Nov. 18, 2016), https://www.vanityfair.com/news/2016/11/obama-slams-facebook-for-outright-lies-online; Ashley May, *How Facebook Plans To Crack Down On Fake News*, USA TODAY (Nov. 19, 2016), https://www.usatoday.com/story/tech/2016/11/19/how-facebook-plans-crack-down-fake-news/94123842/ [https://perma.cc/U2BD-A43D].

[111] Kosoff, *infra* note 118.

[112] *Id.*

[113] Yu-Ru Lin et al., PLoS ONE, Rising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events (2014), https://doi.org/10.1371/journal.pone.0094093 [https://perma.cc/RBQ3-5AK4].

events."[114] Thus, the impact of social media on political events is well within the current considerations of researchers. Further research is needed to determine how to develop resiliency in the face of misinformation.

D. Surveillance Based Influence and Manipulation

Facebook has been involved with several experiments that have caused heated debate within the social science world each because of Facebook altering feeds or information to influence individuals' behaviors. For example, in 2010, Facebook experimenters assigned Facebook users into three categories during the election. One group received a simple information message about the election, one group received no information at all, and one group received information and were given the ability to click an 'I voted' button.[115] The results showed that those who got the informational message voted at the same rate as those who saw no message at all.[116] But those who saw the social message were 2% more likely to click the 'I voted' button and 0.3% more likely to seek information about a polling place than those who received the informational message, and 0.4% more likely to head to the polls than either other group.[117] The researchers estimated this resulted in about 340,000 extra people turning out to vote in the 2010 U.S. congressional elections because of a single election-day Facebook message.[118]

According to noted authority Professor Baldwin, "'Nudging' involves structuring the choices that people make so as to lead them towards particular outcomes."[119] Space constraints do not permit a full exploration of the theory of "nudging,' but one particular type of nudge should draw our immediate attention, the 'Third Degree Nudge,' which involves behavioral manipulation that uses "a message with an emotional power that blocks the

---

[114] *See id.*

[115] Robert M. Bond et al., *A 61-Million-Person Experiment In Social Influence And Political Mobilization*, 489 NATURE 295 (2012), www.nature.com/doifinder/10.1038/nature11421 [https://perma.cc/G523-9RGN].

[116] *See id.*

[117] *See id.*

[118] *See id.* It should be noted, Cameron Marlow, head of Facebook's data-science team and a co-author of the paper, stresses that individuals' identities had been protected in the study. *See id.*

[119] Baldwin, *supra* note 102. There are actually three categories of nudges. 'First Degree Nudges,' such as simple warnings or reminders, are designed to respect the decision-making autonomy of the individual and serve no other purpose than to enhance the individual's reflective decision-making process. A 'Second Degree Nudge' is designed to use 'choice architecture' to build on behavioral limitations in an effort to bias a decision in the desired direction. *See generally*, RICHARD THALER & CASS SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2009); DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011).

consideration of all options."[120] Commentators argue that this type of behavior manipulation directly impacts an individual's ability to act in accordance with her or his own preferences.[121] For example, according to Sarah Buhr at Tech Crunch Apple maps has been directing people searching for abortion clinics to adoption centers since 2011.[122] This type of nudge is not accidental, nor is it unnoticed. Yet, it persists even in the face of confrontation by various groups. We must confront the reality that new behavior patterns on socially significant scales may be driven by the decisions of major marketers and service providers operating with no public mandate.

IV. UNPACKING THE REGULATORY LANDSCAPE

This Part briefly unpacks the regulatory landscape focusing on the role of the U.S. Federal Trade Commission before moving on to discuss global developments in Big Data and data privacy. Subsequently, we pivot to examine policy implications resulting from governance gaps. Due to the fractured governance structure in play, we also rely on the field of polycentric institutional analysis to give shape to both our analysis of the regulatory status quo, and to lay out a path for policymakers going forward.

A. A Polycentric Primer

The field of polycentric (multi-centered) governance is a multi-level, multi-purpose, multi-functional, and multi-sectoral model[123] that has been championed by scholars, including Nobel Laureate Elinor Ostrom and Professor Vincent Ostrom. It challenges orthodoxy by demonstrating the benefits of self-organization, networking regulations "at multiple scales,"[124] and examining the extent to which national and private control can in some cases coexist with communal management.[125] The field also posits that, due

---

[120] *See* Baldwin, *supra* note 119, at 2.

[121] *See id.*

[122] Sarah Buhr, *Apple Maps Has Been Directing People Searching For Abortion Clinics To Adoption Centers Since 2011*, TECH CRUNCH (Jan. 29, 2016), http://tcrn.ch/1KLgfWY [https://perma.cc/TQD3-JYCN]. Interestingly, Apple has known about the 'glitch' since 2011 but has never taken action to fix the problem. *Id.*

[123] Michael D. McGinnis, *An Introduction to IAD and the Language of the Ostrom Workshop: A Simple Guide to a Complex Framework*, 39(1) POL'Y STUD. J. 163, 171–72 (Feb. 2011).

[124] Elinor Ostrom, *Polycentric Systems as One Approach for Solving Collective-Action Problems*, 1 (Ind. Univ. Workshop in Political Theory and Policy Analysis, Working Paper Series No. 08–6, 2008), http://dlc.dlib.indiana.edu/dlc/bitstream/handle/10535/4417/W08-6_Ostrom_DLC.pdf?sequence=1.

[125] For a detailed discussion of early Internet history, *see* KATIE HAFNER & MATTHEW LYON, WHERE WIZARDS STAY UP LATE: THE ORIGINS OF THE INTERNET (1996); Barry M. Leiner et al., *Brief History of the Internet*, INTERNET SOC'Y (Oct. 29, 2017, 4:40 PM), https://www.internetsociety.org/internet/history-internet/brief-history-internet/ https://perma.cc/4DY8-BLR4].

to the existence of free riders in a multipolar world, "a single governmental unit" is often incapable of managing "global collective action problems" such as cyber-attacks.[126] Instead, a polycentric approach recognizes that diverse organizations working at multiple levels can create different types of policies that can increase levels of cooperation and compliance, enhancing "flexibility across issues and adaptability over time."[127] Such an approach, in other words, recognizes both the common but differentiated responsibilities of public- and private-sector stakeholders and the potential for best practices to be identified and spread organically, generating positive network effects that could, in time, result in the emergence of a norm cascade toward greater clarity in Big Data and privacy regulation.[128]

The field of polycentric governance is far from a perfect fit as applied to Big Data and Machine Learning, and the literature itself is incomplete, as Professor Elinor Ostrom would have been the first to admit.[129] However, it can provide insights about useful paths forward to generate norms within fractured governance structures, and is, in fact, already the preferred path forward from leading public- and private-sector stakeholders.[130] This is especially true when these findings are coupled with the wider literature on regimes and institutional analysis, including its focus on property rights and transaction costs, as is discussed further next in the context of applicable federal and international law.[131]

B. The Federal Trade Commission

While many state and federal agencies may be implicated in bringing actions against the activities described in Parts II and III, no agency is arguably more central in the discussion than the Federal Trade Commission (FTC). The FTC is a government agency that is responsible for protecting the public against unfair business practices, including false and misleading

---

[126] Elinor Ostrom, *A Polycentric Approach for Coping with Climate Change,* WORLD BANK 35 (World Bank Policy Research Working Paper No. 5095, 2009).

[127] Robert O. Keohane & David G. Victor, *The Regime Complex for Climate Change,* 9 PERSP. ON POL. 7, 9 (2011); *cf.* Julia Black, *Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes*, 2 REG. & GOVERNANCE 137, 157 (2008) (discussing the legitimacy of polycentric regimes, and arguing that "[a]ll regulatory regimes are polycentric to varying degrees").

[128] *See* Martha Finnemore & Kathryn Sikkink, *International Norm Dynamics and Political Change*, 52 INT'L ORG. 887, 895–98 (1998).

[129] *See* Elinor Ostrom, et al., *Revisiting the Commons: Local Lessons, Global Challenges*, 284 SCI. 282, 282 (1999) (noting that some of her work in the global commons context to "provide starting points for addressing future challenges.").

[130] *See* Nancy Scola, *ICANN Chief: "The Whole World is Watching" the U.S.'s Net Neutrality Debate*, WASH. POST (Oct. 7, 2014).

[131] For more on this topic, *see* LEE J. ALSTON, THRAINN EGGERTSSON, & DOUGLASS C. NORTH, EMPIRICAL STUDIES IN INSTITUTIONAL CHANGE 92 (1996).

advertising.[132] The agency has enforcement authority under its enabling statute, the Federal Trade Commission Act (FTC Act),[133] which is a federal consumer protection law that prohibits unfair or deceptive practices, including offline and online privacy and data security policies.[134] The FTC has used this authority to, among other things, help define private-sector cybersecurity best practices as an example of polycentric norm building efforts,[135] and is also the primary enforcer of the Self-Regulatory Principles for Behavioral Advertising.[136]

To assist companies in managing online consumer based data gathering and use, the FTC issued a report entitled "Big Data: A Tool for Inclusion or Exclusion."[137] To "maximize the benefits [of Big Data] and limit the harms,"[138] the FTC advises a company to consider the following questions:

- How representative is your data set?[139]
- Does your data model account for biases?[140]
- How accurate are your predictions based on Big Data?[141]
- Does your reliance on Big Data raise ethical or fairness concerns?[142]

With this guidance, the FTC seeks to enforce the FTC Act and to encourage businesses to follow the law and principles described above. To date, the FTC has brought many enforcement actions against companies failing to comply with posted privacy policies and for the unauthorized disclosure of personal data.[143] FTC enforcement actions have been in the new

---

[132] *About the FTC*, FED. TRADE COMMISSION, (last updated Oct. 29, 2017, 5:53 PM), https://www.ftc.gov/about-ftc [https://perma.cc/RJY9-MWXU].

[133] 15 U.S.C. §§41–58 (2012).

[134] *Id.*

[135] *See, e.g.*, Lesley Fair, *Wyndham's Settlement with the FTC: What it Means for Businesses – and Consumers,* FED. TRADE COMMISSION (Dec. 9, 2015), https://www.ftc.gov/news-events/blogs/business-blog/2015/12/wyndhams-settlement-ftc-what-it-means-businesses-consumers/ [https://perma.cc/9T5F-F6LX].

[136] DIGITAL ADVERTISING ALLIANCE, *Self-Regulatory Principles for Online Behavioral Advertising* (2009), https://www.iab.com/wp-content/uploads/2015/05/ven-principles-07-01-09.pdf [https://perma.cc/7WVU-FC4Q].

[137] Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues* (Jan 2016).

[138] *See id.* at vi.

[139] *See id.*

[140] *See id.* at iv.

[141] *See id.* at v.

[142] *See id.*

[143] Patricia Bailin, *What FTC Enforcement Actions Teach Us About the Features of Reasonable Privacy and Data Security Practices*, Westin Research Center, *available at* https://iapp.org/media/pdf/resource_center/FTC-WhitePaper_V4.pdf.

more frequently as of late, for example, the 2007 corporate and in-store network exploitation by a hacker at the chain restaurant Dave & Buster's,[144] and the 2006 and 2007 insecure disposal of personal information by CVS pharmacies,[145] among others. Relatively few enforcement actions, though, have been brought under the January 2016 FTC Big Data report; however, the language here is clear, which is that the commission "will continue to monitor areas where Big Data practices"[146] could violate those laws "and will bring enforcement actions where appropriate."[147] To what degree such enforcement actions will continue under the Trump administration remains to be seen.[148]

C. Global Data Laws and Regulations: Non-Discrimination Case Study

Firms operating multinationally, though, are making decisions on Big Data and machine learning in reference to numerous legal regimes outside of the United States. Without going into too much detail, as this article is not set up to be a review of privacy law and regulation,[149] information privacy or data protection laws prohibit the disclosure or misuse of information held on private individuals. More than eighty countries and independent territories have now adopted comprehensive data protection laws, including nearly every country in Europe and many in Latin America and the Caribbean, Asia, and Africa.[150] It is, however, important to note that the U.S. has not adopted a comprehensive information privacy law, instead opting for a sector-specific approach. Examples of this style include the Health Insurance Portability and Accountability Act of 1996 (HIPAA),[151] the Children's Online Privacy Protection Act of 1998 (COPPA),[152] the Fair and Accurate Credit Transactions Act of 2003 (FACTA),[153] the Electronic

---

[144] *Dave & Buster's, Inc*., FTC Docket No. C-4291 (May 20, 2010).

[145] *CVS Caremark Corp*., FTC Docket No. C-4259 (June 18, 2009); *U.S. v. PLS Financial Services, Inc.*, No. 12-CV-08334 (N.D. Ill. Oct. 17, 2012).

[146] Big Data report*, supra* note 137, at v.

[147] *See id.*

[148] *See, e.g.*, Allison Grande*, Trump's FTC May Scale Back On Data Security Enforcement,* LAW 360 (Nov. 10, 2016), https://www.law360.com/articles/861714/trump-s-ftc-may-scale-back-on-data-security-enforcement.

[149] For a 2016 map displaying current regulation, *see* David Banisar, *National Comprehensive Data Protection/Privacy Laws and Bills 2016* (November 28, 2016)., https://ssrn.com/abstract=1951416.

[150] *See* Graham Greenleaf, *Global Data Privacy Laws: 89 Countries, and Accelerating*, Privacy Laws & Business International Report, Issue 115, Special Supplement, February 2012).

[151] *See generally,* Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104–191, 110 Stat. 1936.

[152] *See generally* Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (2012).

[153] *See generally* Fair and Accurate Credit Transactions Act of 2003, 15 U.S.C. §§ 1681–1681x (2012).

Communications Privacy Act (ECPA),[154] and the Family Educational Rights and Privacy Act (FERPA)[155] to name a few.

In general, many of these laws are based on Fair Information Practice and the more comprehensive Organisation for Economic Co-operation and Development (OECD) Principles and the European Union Data Protection Directive. Many would agree, at a minimum, that the basic principles of data protection include:

- For all data collected there should be a stated purpose.
- Information collected by an individual cannot be disclosed to other organizations or individuals unless specifically authorized by law or by consent of the individual
- Records kept on an individual should be accurate and up to date
- To ensure accuracy, there should be mechanisms for individuals to review data about them. This may include periodic reporting
- Data should be deleted when it is no longer needed for the stated purpose
- Transmission of personal information to locations where "equivalent" personal data protection cannot be assured is prohibited
- Some data is too sensitive to be collected, unless there are extreme circumstances that justify such gathering (e.g., sexual orientation, religion)

These agreed upon protections fail to fully envision the ease in which data gathered, even publically available data, can be agglomerated into 'new' data or data that when combined can create a complete picture of a digital personhood. In addition, the protections fail to recognize that data collected can be used in such a manner as to create significant impact on the individual and society as a whole. It is these two unappreciated aspects of data regulation that cause the greatest likelihood of discriminatory impacts, yet; in many ways data regulation fails to anticipate such impacts.

D. Non-Discrimination Legal Regulation

The right to non-discrimination is deeply embedded in the framework that underlies both the United States and the European Union. For example, in the European Union the right can be found in Article 21 of the Charter of Fundamental Rights of the European Union, Article 14 of the European Convention on Human Rights, and in Articles 18-25 of the Treaty on the

[154] *See generally* Electronic Communications Privacy Act, 18 U.S.C. § § 2701–2712 (2000 & Supp 1 2001).

[155] *See generally* Family Educational Rights and Privacy Act, Pub. L. No. 93-380, 88 Stat. 571 (1974) (codified at 20 U.S.C. § 1232g (2012)). The regulations that administer FERPA are incorporated in 34 C.F.R. § 99.

Functioning of the European Union, to name but a few.[156] In a similar manner, the United States has attempted to reduce discrimination by crafting policy. A simple example demonstrates the breadth of the regulation. Within the area of employment, federal law protects individuals from discrimination relating to: age, disability, equal pay/compensation, genetic information, harassment, national origin, pregnancy, race/color, religion, retaliation, sex, and sexual harassment, to name a few.[157] And while each of these areas place people within a protected class, it is important to note that those individuals that fall within a protected class vary based upon the federal law that is implicated. The area is complex, murky, and often difficult to describe. For the purposes of this article, what is most concerning, however, is the means by which an individual demonstrates the existence of a discriminatory practice. Returning to the employment example, in the United States, an employee who believes they have been discriminated against based on their status as a member of a protected class has several types of claims: discriminatory intent/treatment disparate impact,[158] and/or retaliation.[159] And while each of these categories is broad, the evidentiary rule is quite narrow in the area of employment law. For example, to demonstrate the existence of discrimination the employee must produce either direct evidence or circumstantial evidence. And of course, direct evidence—that is, the gathering of statements or other evidence that directly relates to discriminatory practices—is very rare. Thus, employees are left with the circumstantial route.

To demonstrate discriminatory practices, an employee that seeks to demonstrate a Title VII[160] disparate treatment claim must satisfy the

---

[156] Charter of Fundamental Rights of the European Union art. 21, 2000 O.J. (C 364) 1; Convention for the Protection of Human Rights and Fundamental Freedoms art. 14, Nov. 4, 1950, E.T.S. No. 5, 213 U.N.T.S. 221; Consolidated Version of the Treaty on the Functioning of the European Union art. 18-25, 2012 O.J. (C 326) 47.

[157] *See e.g.*, Civil Rights Act of 1964: Title IX of Education Amendments of 1972, 20 U.S.C. §§ 1681–1688 (2012); Age Discrimination Act of 1975, 42 U.S.C. §§ 6101–6107 (2012); Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12101-12213 (2012).

[158] A disparate impact claim is a type of discrimination based on the effect of an employment policy, rule, or practice rather than the intent behind it. *See, e.g.*, Griggs v. Duke Power Co., 401 U.S. 424, 430 (1971) ("Under [Title VII], practices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices.").

[159] Equal Employment Opportunity (EEO) laws prohibit retaliation and related conduct. *See, e.g.*, Title VII of the Civil Rights Act of 1964 (Title VII), 42 U.S.C. § 2000e-2; Age Discrimination in Employment Act (ADEA), 29 U.S.C. § 623; Title V of the Americans with Disabilities Act (ADA), 42 U.S.C. § 12112; Section 501 of the Rehabilitation Act (Rehabilitation Act), 29 U.S.C. § 791; Equal Pay Act (EPA), 29 U.S.C. § 215; Title II of the Genetic Information Nondiscrimination Act (GINA), 42 U.S.C. § 2000ff-1.

[160] Title VII prohibits discrimination by covered employers on the basis of race, color, religion, sex or national origin. *See* 42 U.S.C. § 2000e-2.

McDonnell-Douglas burden-shifting framework.[161] Named for a famous U.S. Supreme Court decision, an employee must be able to answer "yes" to the following four questions:

- Are you a member of a protected class?
- Were you qualified for your position?
- Did your employer take adverse action against you?
- Were you replaced by a person who is not in your protected class (or, in the case of age discrimination, someone substantially younger than you)?

While these may seem to be simple straight forward questions, the ability to demonstrate discrimination in these settings has proven very difficult.[162] Thus, the framework shifts the burden of production to the employer to "articulate some legitimate, nondiscriminatory reason for the employee's rejection."[163] It is important to note, employment law within the U.S. has developed a mechanism to handle situations in which it would be difficult (or impossible) for an individual to gather the evidence necessary to

---

[161] McDonnell Douglas v. Green 411 U.S. 792 (1973) and Texas Dept. of Community Affairs v. Burdine, 450 U.S. 248 (1981). The other common test is the Price Waterhouse "mixed motive" framework.

[162] Of course, this is one example—in fact, the law recognizes that persons can be discriminated against even if they were not replaced by someone outside of the protected class. Several authorities within the area have suggested employees consider of the following criteria as well:

- Were you treated differently than a similarly situated person who is not in your protected class?
- Did managers or supervisors regularly make rude or derogatory comments directed at your protected class status or at all members of your class and related to work? For example, "Women don't belong on a construction site" or "Older employees are set in their ways and make terrible managers."
- Are the circumstances of your treatment so unusual, egregious, unjust, or severe as to suggest discrimination?
- Does your employer have a history of showing bias toward persons in your protected class?
- Are there noticeably few employees of your protected class at your workplace?
- Have you noticed that other employees of your protected class seem to be singled out for adverse treatment or are put in dead-end jobs?
- Have you heard other employees in your protected class complain about discrimination, particularly by the supervisor or manager who took the adverse action against you?
- Are there statistics that show favoritism towards or bias against any group?
- Did your employer violate well-established company policy in the way it treated you?
- Did your employer retain less qualified, non-protected employees in the same job?

PAUL H. TOBIAS & SUSAN SAUTER, JOB RIGHTS AND SURVIVAL STRATEGIES; A HANDBOOK FOR TERMINATED EMPLOYEES, Jist Publishing (January 1997).

[163] McDonnell Douglas v. Green, 411 U.S. 792 at 802 (1973).

prove their claim. Such burden shifting is somewhat common in situations where such difficulties arise. However, even an area of law that is comfortable with information gathering difficulties will still be unable to accommodate the difficulties created through the use of machine learning. Remember, in a machine-learning decision-making world, even those well versed in machine learning may be unable to roll back time and decipher black box learning.

The inability to peer inside the black box is not a minor inconvenience, and as can be seen within the area of employment law, even areas accustomed to the black box dilemma will have difficulty handling such a problem. Of course, employment is but one example, consider many of the cases mentioned above—policing, insurance, credit worthiness. The widespread use of machine learning will lead to black box based decision making in many areas of our daily lives, with little to no mechanism of examination or ability to demonstrate discrimination. Consider the case of Google targeted ads mentioned above. As you may recall, Carnegie Mellon researchers built a tool to simulate Google users that started with no search history and then visited employment websites.[164] Later, on a third-party news site, Google showed an ad for a career coaching service advertising "$200k+" executive positions 1,852 times to men and 318 times to women.[165] While Google will not disclose the manner in which it targets ads,[166] the outcome is stark. Keep in mind, targeting ads is legal, discriminating on the basis of gender is not. But, how is one to know which action is behind such an outcome?

At least partially because of the inability to peer inside the box after a machine learning based decision, the European Union has taken a more regulatory approach when discussing machine learning. For example, starting in 2018, EU citizens will be entitled to know how an EU institution arrived at a conclusion—even if machine learning and a black box was involved.[167] As University of Oxford researcher Bryce Goodman explains, the new data protection law entitled the General Data Protection Regulation (GDPR) is effectively a "right to an explanation"[168] for decisions.[169] In fact, the law does more; it also bans decisions "based solely on automated

[164] Claire Cain Miller, *When Algorithms Discriminate*, N.Y. Times (July 13, 2015), https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html.

[165] *See id.*

[166] *See id.*

[167] Bryce Goodman & Seth Flaxman, *EU Regulations On Algorithmic Decision-Making and a 'Right to Explanation,'* 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, *available at* https://arxiv.org/abs/1606.08813.

[168] *Id.* at 6.

[169] *Id.*

processing, including profiling, which produces an adverse legal effect concerning the data subject or 'significantly affects' him or her."[170] Unfortunately, as can be seen from the cited language, the use of terms such as 'solely' will leave large loopholes in the discussion about the use—or requirements to use—human beings in the decision making processes. Moreover, it is possible that an overly expansive reading of the language will lead to a stifling of innovation in positive uses of machine learning. What is clear, however, is that the attempted regulation demonstrates a recognition of the growing ethical difficulties with black-box HAL-based decision making.

E. Implications for Policymakers and Managers

In addition to the FCC recommendations (and others), it is time for policymakers to think beyond privacy, specifically to impact; a point that the European Union is considering as a linchpin of legal analysis. Like the EU, we would like to suggest that U.S. regulation within this area must begin to consider modern technology and thus begin to focus on the impact of the use of information collected. The above examples highlight the need to begin to consider polycentric interventions (including industry norms) that are designed to contemplate the negative impacts created from the amalgamation of large data sets, even if that information is all considered public information or has been gathered with consent.

It is important to recognize one key point in the above suggestion—the information being discussed in many of these examples is either public, or information that has been gathered and used with consent. We argue that consent should not be treated the same in the online environment.[171] The nature of disclosures, informed individuals, and consent has changed with the ability of entities to push out large amounts of information and to obtain consent with a mere click of the button; a button that often stands as the only barrier before the consumer can obtain a coveted item.[172] The laws that apply to consent must be reconsidered in light of the new research that demonstrates that consumers are presented with too much information, often in unmanageable formats with little desire or need to read or understand terms.[173] Consent cannot cure this new world of disparate impact; neither can

[170] *Id.* at 2.

[171] Anjanette H. Raymond, *Yeah, But Did You See the Gorilla? Creating and Protecting an 'Informed' Consumer In Cross-Border Online Dispute Resolution*, 19 HARV. NEGOT. L. REV. 129 (Spring 2014).

[172] *See id.* at 146.

[173] *See id.*

transparency.[174] We have to begin to consider the impact of the use of information and design regulation that is not overcome by a simple disclose and a mere click.

Moreover, while issues of consent must be addressed, the resolution of this issue does not eliminate—or even dent—the concerns that arise in relation to "digital personhood." By digital personhood, we mean the profile of each of us that is created through the use of public information.[175] As discussed above, we all share vast amounts of information; none of which is protected under the current regulatory approaches to data protection and/or privacy protections. What then must be immediately done to hold algorithms accountable?

First, algorithms used within an area of potential discriminatory impact, regardless of the source of the information being used, must be held to social and statistical science accountability standards. This requires three things: (1) accountability, (2) auditability, and (3) replication. The social science profession has long subscribed to the ethics of applied social research, which includes the principle of *voluntary participation* and *informed consent*. Moreover, ethical standards require considerations of *risk of harm*, which would include potential discriminatory impacts. Of course, each of these principles demand a level of accountability for the designer and the user of the information—with the power to adjust and respond to issues that arise. And finally, the power of auditability—and replication—are essential attributes with social sciences as it provides scientists the ability to monitor, check and criticize both the design, the outcomes and the impact of the uses of the information and outcomes.

Second, all decisions produced through a process that includes an algorithm must be capable of explanation of outcome and the process that occurred, including any decisions that were the product of a weighted decision making process. It is important to understand not all attributes that can appear at first blush as discriminatory actually have discriminatory impacts. For example, being a black male increases your risk of inheriting sickle cell anemia.[176] Clearly these attributes built into some algorithms have discriminatory impact, but in a medical diagnosis context, the impact must be balanced against the medical needs of the individual and society's desire to improve the health of its citizenry. Algorithms are taught to weigh certain

---

[174] *See* Fred H. Cate, *Protecting Privacy in Health Research: The Limits of Individual Choice,* 98 CAL. L.REV. 1765, 1769 (2010)

[175] *See* Roger Clarke, *The Digital Persona and Its Application to Data Surveillance*, THE INFO. SOC'Y 77 (1994).

[176] *See* The National Heart, Lung, and Blood Institute, *Sickle Cell Disease*, available at https://www.nhlbi.nih.gov/health-topics/sickle-cell-disease [https://perma.cc/9G6M-FDKD].

attributes by the programmers and the learning process. These weightings, attributes and other factors that impact the outcomes of the algorithm must be capable of explanation and challenge, especially within areas of potential discriminatory impact.

Third, all outcomes of an algorithm based process must be capable of challenge by individuals and institutions. Mistakes occur even in machine-based decision making, from data entry mistakes to statistical errors. Individuals impacted by algorithm based decision making need the ability to challenge incorrect information and assumptions within the environment. Allowing this type of correction is not merely important for the individuals, but also allows those who design the algorithm to understand the nature of the errors and correct for errors in the future.

The absence of U.S. law governing the amalgamation of information and the unwavering belief that transparency is the answer to all that ails online information gathering, leads us to argue that ethical considerations must begin to be a larger part of the discussion surrounding the manner in which information should be used and regulated.

V. AN ETHICAL MODEL FOR EXAMINATION OF NEW DILEMMAS

While each of the above mentioned issues have probably given readers more than a moment's pause, Silicon Valley is best known for pushing out products and not for considering the societal or ethical considerations of the use of their products.[177] Consider the case of the Internet of Things (IoT). Many commentators,[178] including two of this paper's authors,[179] have called for a greater emphasis to be placed on security within the IoT design framework.[180] Yet, it took an October 2016 attack in which "smart" home devices were used as a major component in a DDoS attack to demonstrate the true need of heightened built-in security on internet connected home devices.[181]

Silicon Valley is in the business of delivering products that people want to buy and use, it is a market based economy at its best. In fact, it seems that

[177] *See, e.g.*, Om Malik, *Silicon Valley Has an Empathy Vacuum*, THE NEW YORKER, Nov. 28, 2016.

[178] *See e.g.,* Press Release, Gartner Newsroom, Gartner Says IoT Security Requirements Will Reshape and Expand Over Half of Global Enterprise IT Security Programs by 2020, Gartner Newroom (May 1, 2014), http://www.gartner.com/newsroom/id/2727017.

[179] *See, e.g.*, Scott J. Shackelford et al., *When Toasters Attack: A Polycentric Approach to Enhancing the "Security of Things,*" U. ILL. L. REV. 415 (2017).

[180] Amir Nasr, *HECC reps urge FTC to make sure IoT developers create with security in mind*, Daily Dashboard, (Nov. 4, 2016), https://iapp.org/news/a/hecc-reps-urge-ftc-to-make-sure-iot-developers-create-with-security-in-mind/.

[181] BBC Author*, 'Smart' home devices used as weapons in website attack*, BBC News (Oct. 22, 2016), http://www.bbc.com/news/technology-37738823.

those in the best position to consider and debate such an issue, those within the group of the "data-driven oligarchy like Facebook, Google, Amazon, or Uber"[182] wish to wash their hands of the real-world filter bubbles in which "people become numbers, algorithms become the rules, and reality becomes what the data says."[183] In response, *New Yorker* author Om Malik notes:

> Facebook's blunders are a reminder that it is time for the company to think not just about fractional-attention addiction and growth but also to remember that the growth affects real people, for good and bad. It is not just Facebook. It is time for our industry to pause and take a moment to think: as technology finds its way into our daily existence in new and previously unimagined ways, we need to learn about those who are threatened by it. Empathy is not a buzzword but something to be practiced.[184]

With the absence of regulation and a widespread cultural attitude within the design industry of ignoring the wider context of technology, how can we ever begin to consider the larger ethical questions surrounding the use and impact of technology in our lives? In *Media Technologies*,[185] Tarleton Gillespie outlines six "categories of ethical concern": (1) Patterns of Inclusion (how data is selected for indexing); (2) Cycles of Anticipation (the role of hypothetical models of consumer behavior in shaping algorithm design); (3) Evaluation of Relevance (the sorting protocols of the algorithm itself); (4) Promise of Objectivity (the often problematic rhetoric that surrounds innovations in Big Data and AI); (5) Entanglement with Practice (new technological phenomena considered within the full context of actual implementation); (6) Production of Publics (basically a higher-level view of how 2, 4, and 5 interact to create new human social groups). [186]

To consider an example application of Gillespie's Model, let's return to Bostrom and Yudkowsky's scenario of mortgages issued by a credit worthiness algorithm.[187] The bank in the mortgage example may be completely honest that it tried to institute a colorblind approval process.[188] In this case, the problem lies in bias fossilized in the data set. The overwhelming majority of the history of banking has been shaped by racist discrimination, up to and through the 2008 housing market crash[189] – and the

---

[182] Malik, *supra* note 177, at 2.

[183] *Id.* at 4.

[184] *Id.* at 5.

[185] *See* Tarleton Gillespie, Pablo J. Boczkowski and Kirsten A. Foot (eds.), MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY, MIT Press, (2014).

[186] *Id.*

[187] Bostrom, *supra* note 18, at 316.

[188] *Id.*

[189] Ta-Nehisi Coates, *The Case for Reparations*, THE ATLANTIC (June 2014), https://www.theatlantic.com/magazine/archive/2014/06/the-case-for-reparations/361631/.

past is the source of the data set. If a supervised algorithm has been trained on historical data, factors closely correlated with race are also likely to be correlated with acceptance or denial, even if no explicit "race" data points are included. As for an unsupervised algorithm, if it picked out and re-applied the pattern of discrimination on its own it would actually be doing its job – as a pattern detector – exceedingly well.

As Bostrom and Yudkowsky explain:

> If the machine learning algorithm is based on a complicated neural network, or a genetic algorithm produced by directed evolution, then it may prove nearly impossible to understand why, or even how, the algorithm is judging applicants based on their race. On the other hand, a machine learner based on decision trees or Bayesian networks is much more transparent to programmer inspection.[190]

Bostrom and Yudkowsky's discussion cautions us that while a lay person, and even a policymaker, might initially believe that you can assign blame to the programmer for setting up "racist parameters," depending on the type of machine learning being used, this might be both unfair and futile.[191] We need to examine the data set, the institutional context, and the mediating role of the machine, with the kind of critical literacy we apply in other cases of institutional innovation.

In Gillespie's terms,[192] the Bostrom and Yudkowsky's scenario asks us to consider both the patterns of inclusion that shape the data set, and backtracking to the evaluation of relevance produced by machine learning, by looking first for red flags in the outputs.[193] What we learn in this example reflects not just on the quality of the machine learning, but goes all the way back, prior to the algorithms, prior to the data set, to reflect on problems in the institution of banking itself.

This points to Gillespie's fifth area of ethical concern: entanglement with practice.[194] We tend to think of new technology as a source of new knowledge production, because that is the most exciting outcome – and can be wildly useful.[195] But reaffirming common sense or current practice is often useful as well, especially in institutional settings. Perhaps the outcome in Bostrom and Yudkowsky's example should not surprise anyone. Keep in mind that a "successful" machine learning algorithm is almost always going to be defined as such based on being *useful* or at least *usable* in a specific

[190] Bostrom, *supra* note 18, at 316.
[191] *Id.* at 317.
[192] Gillespie, *supra* note 185, at 168–9.
[193] Bostrom, *supra* note 18, at 316.
[194] Gillespie, *supra* note 185, at 168.
[195] *Id.* at 190.

human context. It is successful if it is able to process available data in a way that either provides its human users with something they can apply for decision-making within pre-existing systems, or even goes ahead and makes the decisions that will be legible and minimally disruptive in the context of existing human institutional frameworks.

What remains troubling is the imaginary bank's response, believing that their algorithm, simply by *being* an algorithm, has the power to erase the history of their racist and predatory lending policies. This points to Gillespie's fourth area of concern: the rhetorical promise of objectivity is taking at least as large a role in the story as the program's actual computing capacities.[196] In fact, it is a successful example of machine learning *precisely because* it has learned on its own to apply to the future the racist and predatory policies it learned from the past. The promise of objectivity becomes a sort of invisibility cloak, wrapping up pre-existing social bias and re-engraining it, causing moral, and possibly legal agency to quietly disappear in the process.

The ethical discussion, framed around Gillespie's Model[197] allows us to discuss new issues as they arise within the field, and to better inform corporate decision-making about the use of collected data from machine learning. It is only with a continuing robust discussion as new uses of data and ever increasing amalgamation of data can we begin to design regulation that matches the expectations of society.

CONCLUSION

Big Data is already being used in many settings with positive outcomes. For example, the FTC report referenced above notes that use of Big Data (and analytic techniques) can identify students who are at risk of dropping out and in need of early intervention strategies,[198] examine academic disciplinary actions that are disproportionately impacting certain populations,[199] provide access to credit to a larger segment of the population through a more robust credit ranking system,[200] provide individualized

[196] *Id.* at 179.

[197] *Id.* at 168.

[198] *See, e.g.*, F.T.C., Transcript of Big Data: A Tool for Inclusion or Exclusion?, 84–85 (Sept. 15, 2014).

[199] *See id.* at 49–51.

[200] *See, e.g.*, Gene Gsell, Transcript of Big Data: A Tool for Inclusion or Exclusion?, 49–51 (Sept. 15, 2014).

healthcare,[201] and to provide healthcare into underserved communities.[202] But the report also notes that there have also been instances of Big Data misuse. For example, the FTC took action against one credit card company because it failed to disclose its practice of rating consumers as having a greater credit risk because they used their cards to pay for marriage counseling, therapy, or tire-repair services.[203] And, at least one example exists of online companies charging consumers in different zip codes different prices for standard office products.[204]

Given that comprehensive regulation is unlikely in this area for the foreseeable future outside of the EU context, firms will need to make the proactive decision of how they should ethically use the massive amounts of data being generated through machine learning processes. We argue that the most productive path forward is for corporate decision-makers is to leverage the power of polycentricity and proactively promote ethical Big Data norm building. This could include incorporating best practices from related contexts, such as cybersecurity, which has seen the rise of Information Sharing and Analysis Centers (ISACs) and Organizations (ISAOs) to share cyber threat data and best practices between members. The same could be done in the Big Data context with regards to privacy threats and best practices.[205] Without such polycentric action, the warnings of science fiction could soon become reality.

---

[201] *See, e.g.*, Shannon Pettypiece & Jordan Robertson, *Hospitals Are Mining Patients' Credit Card Data to Predict Who Will Get Sick*, BLOOMBERG (July 3, 2014); David Shaywitz, *New Diabetes Study Shows How Big Data Might Drive Precision Medicine*, FORBES (Oct. 30, 2015).

[202] *See, e.g.,* F.T.C., *supra*, note 197 at 84.

[203] F.T.C., *supra* note 137, at 9 (citing FTC v. CompuCredit Corp., No. 1:08-cv-1976-BBM-RGV (N.D. Ga. June 10, 2008)).

[204] *Id*. at 11 (citing Lauren Kirchner, *When Big Data Becomes Bad Data*, ProPublica (Sept. 2, 2015) (finding that areas with high density of Asian residents are often charged more for the Princeton Review's online SAT tutoring).

[205] *See, e.g.*, National Council of ISACs, https://www.nationalisacs.org/member-isacs (last visited Jan. 30, 2017).